CONSTRUCTION OF AN INTEGRATED MEASURE OF THE BURDEN OF DISEASE

Final Report

Contract Number: HHS-100-97-0005

Prepared for:

Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services

Project Officer: Henry Krakauer, M.D., Ph.D.

Prepared by:

The Lewin Group

David Stapleton, Ph.D.
Selen Opcin

July 13, 2001

CONSTRUCTION OF AN INTEGRATED MEASURE OF THE BURDEN OF DISEASE

Final Report

Contract Number: HHS-100-97-0005

Prepared for:

Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services

Project Officer: Henry Krakauer, M.D., Ph.D.

Prepared by:

The Lewin Group

David Stapleton, Ph.D.
Selen Opcin

July 13, 2001

TABLE OF CONTENTS

I. Introduction

A. Background

1. The Burden of Disease

There has been a long-standing social and scientific interest in measuring the "burden of disease" – the cost to individuals and society at large of illnesses, injuries, etc. In developed countries, interest in recent decades has in part been driven by the increasing burden of one component of the burden of disease, namely resources used explicitly to provide health care services – "health care costs" in conventional terminology. But this is just one component of the burden of disease. The burden also includes reduction in the quality of life and in productivity that may be associated with the morbidity caused by the disease or injury – in the extreme, the value of life lost due to premature mortality. In many instances, health care costs are just a small share of the burden of a disease – perhaps more so in developing countries, where expenditures on health care services are relatively low.

A main reason for measuring the burden of disease is to compare the burden of disease under various public health policies, health care services, treatment protocols, and health care systems (e.g., managed care vs. fee for service, or the military health system vs. the civilian health system, or national health insurance vs. a privatized system, or one country's health care system vs. another's). Developing public policies aimed at reducing the burden of disease is a significant challenge.

One important reason for the challenge is the fact that the burden of disease has multiple components (medical costs, reduced quality of life, reduced longevity, productivity loss, etc.). There are often trade-offs between these components. The rate at which one component can be traded for another may differ across policy regimes, treatment protocols, systems, etc. Observed outcomes for these components reflect the available technology, the efficiency with which resources are used in implementing the technology (economic efficiency), preferences of patients, providers, and others, and population risk factors. Furthermore, choices actually made will likely reflect the preferences of consumers and their providers as well as the economic and social environments in which they operate. We cannot claim, for instance, that one delivery system is preferred to another simply because mortality is lower without considering differences in morbidity, quality of life and health care costs.

2. Conventional Methods for Assessing the Burden of Disease

A variety of methods have been used in the literature to assign dollar weights to the various components of the burden of disease. These methods focus on measuring individual or social preferences for the outcome components. Cost-effectiveness analysis – calculation of health care costs per physical unit of "health output" – where health output is measured in a variety of ways, often using fixed weights to "add up" multiple outcomes – is one broad approach. Cost-effectiveness analysis is used to assess whether one system or treatment is less expensive per unit of health output than another, but does not consider whether the value of the benefit from a single system or treatment is at least as great as the cost.

Comparing the values of health outcomes to the cost of producing them is the province of cost-benefit analysis, where the multiple dimensions of health outcomes are themselves valued and compared to health care costs. As with cost-effectiveness analysis, weights are usually applied to the various components of health outcomes to aggregate them into a single outcome measure. The difference in this case is that each weight is intended to represent the value of the corresponding outcome. The aggregate measure is intended to represent the combined values of these outcomes, which can be compared to the costs of producing them.

Both of these approaches to burden of disease comparisons break down because of disagreements over the weights to be applied to the health outcome components. These disagreements, although often reasonable, can lead to very large variations in the final comparisons.

B. Project Objectives

The Office of the Assistant Secretary for Planning and Evaluation in the Department of Health and Human Services has contracted with The Lewin Group to critically assess the treatment of multi-attribute health outcomes in conventional methods for measuring the burden of disease and to explore possible alternatives. This report is the final report of the project. It discusses our review and critique of the conventional methodologies and presents an alternative approach that we have applied to a sample of patients treated in the military health system.

C. Overview of the Report

Section II of the report reviews the conventional approaches taken and discusses our critique of these approaches. ***Section III*** presents hedonic price functions, an alternative approach that has not been considered in the health care literature, but that has been applied to goods and services that have multiple attributes. ***Section IV*** discusses the estimation techniques we use to implement hedonic price functions. ***Section V*** discusses the data used. ***Section VI*** presents the estimated trade-offs between cost and morbidity. ***Section VII*** presents the differences between the estimated trade-offs between cost and morbidity across hospital types. ***Section VIII*** concludes the paper.

II. Conventional Approaches

A. Introduction

The purpose of this section is to understand how the health care literature treats the multi-attribute outcomes of the burden of disease. We will study the already available univariate measures that integrate multiple components of the burden of disease and assess how the weights assigned to each component are determined. We will then examine whose valuations these weights represent and if they approximate currently accepted tradeoffs.

Early measures of the burden of disease concentrated on mortality. Mortality rate is a good indicator of the burden of disease if the time from onset to death is short, as is the case with infectious diseases. In the twentieth century, however, the major causes of death have shifted from infectious to chronic diseases (Stoto, 1992). Individuals live longer with chronic

conditions, often in poor health. The prevalence of chronic conditions has raised the need to measure the quality of life as well as the quantity. This section summarizes measures developed in the health care literature that integrate the quality and quantity of life.

B. Cost-Effectiveness Analysis

Measures that represent years of life weighted or adjusted by the quality of those years have been developed to conduct cost-effectiveness (or cost-utility) analysis. Cost-effectiveness analysis compares the cost of interventions to the health improvement from the same interventions. The most commonly used measure for health improvement is quality adjusted life years (QALY). Healthy years equivalents (HYE) has been proposed as an alternative. These techniques combine health-related quality of life with duration of life to construct a single measure of health improvement that is comparable across interventions. Both measures represent years of life weighted, or adjusted, by the quality of life, although they differ in the way the weights are constructed. Different measures to construct weights imply different approaches to assess the tradeoffs between quality and quantity of life.

Cost-effectiveness analysis calculates cost per unit of health improvement gained. Health improvement and cost are measured separately, with cost in the numerator and health improvement in the denominator. As a result, the measures developed integrate quality and quantity of life after an intervention; they do not include the cost of the intervention.

1. Quality Adjusted Life Years

Each health state from intervention to the end of life is assigned a numerical weight. Weights are added over an individual's lifetime to calculate QALY. Weights are between zero and one, where zero represents death and one represents perfect health. Weights reflect the relative desirability of the health state. Weights can be based on clinical indices or individuals' preferences. Weights must have interval scale properties, which means that the improvement from 0.1 to 0.2 is equivalent to the improvement from 0.7 to 0.8. This is necessary for health improvements across interventions to be comparable.

a. Non-Preference Based Weights

Non-preference based methods choose a clinical index that summarizes the health-related quality of life. Equal scores are assigned to individual components of the health index and summed. Major components of health-related quality of life are physical function, impairment, psychological function, social function, and health perception. The following is a list of the most commonly used health-related quality of life indices and the areas they cover (Goodman, 1998).

- Sickness Impact Profile (SIP): body care and movement, ambulation, mobility, sleep and rest, home management, recreation and past times, emotional behavior, alertness behavior, communication, social interaction, work, eating.

- Quality of Well Being Scale (QWB): symptom-problem complex, mobility, physical activity, social activity.

- Short Form 36 (SF-36): physical functioning, role limitations due to physical problems, social functioning, bodily pain, general mental health, role limitations due to emotional problems, vitality, general health perceptions.

- EuroQol Descriptive System (EuroQol): mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.

- Activities of Daily Living (ADL): bathing, dressing, toileting, mobility, continence, eating.

- Health Utilities Index (HUI): physical function, role function, social-emotional function, and health problems.

Non-preference based weights do not reflect individual or population preferences. Tradeoffs between individual components are arbitrarily set to be equal. While these measures are informative in a clinical sense, they are not informative about desired or actual tradeoffs in health-related quality of life.

b. Preference Based Weights

Preference based methods assign scores to health states by collecting information about individual preferences. These methods derive from utility theory or psychophysical approaches. Preference based weights can use primary or secondary data. For primary data collection, focus groups are asked to rate health states directly through a variety of methods such as the standard gamble or time tradeoff (see below). Focus groups can include patients, health professionals (who have knowledge of the condition) or individuals who have been described a condition through a scenario.

The alternative is to use preference data that has been collected previously. Patients (or health professionals) are asked to describe a health state in terms of a classification scheme (health-related quality of life index). Weights are constructed using preference ratings collected previously. Three widely used data sets, based on the classification schemes described above, are available (Drummond et al, 1997). The Quality of Well Being scale uses category scaling on a random sample of individuals from San Diego, California. EuroQol uses time tradeoff on a random sample of adults in England. The Health Utilities Index uses time tradeoff on a random sample of parents in Hamilton, Ontario. While these data sets are representative for the communities sampled, their representativeness for the US has been questioned.

(1) Utility Approaches to Elicit Preferences

Utility approaches are based on expected utility theory. They measure the desirability of a health outcome (hence its preference based weight) using von-Neumann Morgenstern utilities. They are structured so that they directly produce preference-based weights with interval scale properties. They include standard gamble and time tradeoff techniques (Gold et al, 1996).

- Standard gamble: the respondent is asked to choose between continued life in the health state and a gamble between life in perfect health with probability p or immediate death with

probability 1-p. The probability of life is varied until the respondent is indifferent between the two choices. Recall that perfect health has utility one and death has utility zero. The health state has utility p*, where p* is the probability of indifference. The standard gamble approach involves a choice between a certain outcome and an uncertain outcome. As a result, it captures the individual's risk attitude (most respondents will not like uncertainty), which is a confounding factor in the valuation of health states.

- Time tradeoff: the respondent is asked to choose between continued life in the health state for t years or life in perfect health for x years (where x<t). The value of x is varied until the respondent is indifferent between the two choices. The health state has utility x*/t. The time tradeoff approach involves a choice between two certain outcomes. Hence, it does not capture the individual's risk attitude. It does capture time preference: since years of life given up are at the end of life, they may be valued less because they are in the future.

(2) Psychophysical Approaches to Elicit Preferences

The most widely used psychophysical approaches include the rating scale, category scale, and visual analogue scale (Drummond et al, 1997). These methods ask the respondent to rate directly each health state according to its relative desirability. The intervals between ratings need to correspond to differences in preference perceived by the respondent.

- Rating scale: the respondent is asked to assign a number between 0 and 100 to each health state.

- Category scale: the respondent is asked to assign a discrete point between 0 and 1 (e.g. 0.1, 0.2, 0.3) to each health state.

- Visual analogue scale: the respondent is asked to place each health state between two anchor states, life and death, with no interval marks in between.

Empirical work has shown that people have difficulty assigning a number to health states. There is also no theoretical link between expected utility theory and rating scales. It has been shown that rating scales do not possess interval scale properties. As a result, although easier to administer, psychophysical approaches are less reliable than utility approaches.

2. Healthy Years Equivalents

HYE measures preferences over lifetime health paths instead of over discrete health states like QALY. There are two steps to calculating HYE (Drummond et al, 1997). In step one, the respondent is asked to choose between a lifetime health path and a gamble between life in perfect health with probability p or immediate death with probability 1-p. The probability of life is varied until the respondent is indifferent between the two choices. Step one is equivalent to the standard gamble discussed above, except the certain outcome can include different health states. In the QALY calculation, the certain outcome is the same health state for life. In step two, the respondent is asked to choose between a gamble between life in perfect health for t years with probability p* (where p* is the indifference probability obtained from step 1) or immediate death with probability 1-p* and perfect health for x years (where x<t). x is varied

until the respondent is indifferent between the two choices. The lifetime health path has utility x*/t. Except for measuring utility over a lifetime health path instead of a health state, this measure is a more complicated way of getting at time tradeoff.

3. Disability Adjusted Life Years

Disability adjusted life years were developed as a metric to measure the global burden of disease and disability by a joint effort between the World Health Organization and the World Bank. They represent years of life lost due to premature death or years lived with disability. Years lived with disability are adjusted for the severity of disability. This metric is used to create a data set of the severity of disability caused by several hundred diseases. The weights are based on the preferences of a community of health professionals.

An iterative approach is followed to derive the weights (Murray, 1996). The person tradeoff method is used to derive disability weights for 22 indicator conditions. These weights are used to define seven classes of disability ranging from near perfect health to near death. The rating scale method is used to place each one of several hundred conditions in one of the seven disability classes. The mean disability weight in each disability class is assigned to all conditions that fall within that class.

Disability weights for the indicator conditions are derived in two steps. In step one, two forms of person tradeoff (PTO) were used. In PTO1, the respondent is asked to choose between a life extension for 1,000 healthy individuals and a life extension for 1,000+t individuals in a given health state. The value of t is varied until the respondent is indifferent between the two choices. In PTO2, the respondent is asked to choose between raising the quality of life to perfect health for 1,000+s individuals in a given health state or a life extension for 1,000 healthy individuals, and s is varied until the respondent is indifferent between the two choices. Then, t and s are compared and the respondent is asked to revise his answers if they are not internally consistent. This process is repeated for 22 indicator conditions. In step two, the respondent is asked to rank the 22 indicator conditions according to severity of disability. This ordinal ranking is compared with the cardinal weights assigned to the indicators. If they are not internally consistent, the respondent is asked to revise the weights.

4. Discussion

Cost-effectiveness analysis uses weighted life years to integrate quality and quantity of life to construct a univariate measure of health status. This measure includes disability, morbidity, mortality, and psychological perception; however, it does not include cost. By definition, cost-effectiveness analysis has cost in the numerator and adjusted life years in the denominator, so cost and health improvement is measured separately.

The weights used in weighted life years implicitly place a value on tradeoffs between health-related quality of life and duration of life. Health-related quality of life is represented in an all-in-one measure; the respondent is asked to evaluate all aspects of the health state and provide an (implicit or explicit) weight in terms of duration of life. Rates at which respondents are willing to trade between different aspects of the quality of life are not revealed automatically, although it is possible to infer willingness to trade by repeating the preference eliciting

technique for different conditions and comparing the weights. This is the approach taken to collect data on weights for all components of a classification scheme. By comparing weighted life years associated with each condition, willingness to trade between health conditions can be calculated indirectly.

All methods used in cost-effectiveness analysis to elicit preference based weights rely on hypothetical situations. They are based on what-if scenarios and not real choices. Respondents are asked to choose without having to follow through with a treatment or a payment. As a result, it is impossible to validate the responses or the tradeoffs derived from those responses. It is unclear whether people would behave the same way if they were faced with a real situation.

C. Cost-Benefit Analysis

Cost-benefit and cost-effectiveness analyses both compare the cost of an intervention to the health improvement from an intervention. The difference between the two methods lies in the way health improvements are measured. In cost-benefit analysis, health improvements are measured in monetary terms. Three approaches are used to calculate the monetary value of health improvements: human capital, revealed preference, and contingent valuation (Drummond et al, 1997). These approaches are discussed below.

1. Human Capital Approach

The human capital approach views health care as an investment in human capital. The value of a health improvement is equal to the increased earnings potential it generates. The benefit of an intervention is valued at the present value of future earnings, using market wages. The human capital approach has several drawbacks. It does not value health improvements for retirees or for the unemployed, since they do not have a market wage. Further, in many situations consumer willingness to pay for an intervention may be much greater than the consumer's wage rate because of quality of life effects that go beyond market productivity. One might also argue the opposite, however. Because most workers have sick leave or disability that reduce their cost from lost work time, the value they place on lost work time may be substantially less than their wage. The value of their productivity to society is larger, so their wage might be considered to reflect "societal" willingness to pay for the individual's productivity.

2. Revealed Preference Approach

The revealed preference approach measures the value of a health improvement using market data. For example, the willingness to avoid a health risk can be measured by comparing the wages paid at jobs with and without the health risk. The value of life can be measured by comparing the price of safe versus unsafe cars and the probability of death in an accident. The price difference reflects the market's willingness to pay for the reduction in probability of the health risk (or death). The revealed preference approach is self-validating; valuations are based on behavior, not answers to hypothetical questions. The biggest drawback of the revealed preference approach is that it can only value health events for which data are available. Further, extrapolation of the willingness to pay to avoid an adverse outcome from situations in which the probability of the adverse outcome is extremely small (e.g., death from an automobile

accident in a one-year period) is problematic because individuals may not understand or fully appreciate the risk involved.

3. Contingent Valuation Approach

The contingent valuation approach asks respondents directly how much they are willing to pay to avoid certain health risks. This approach is similar to methods used to elicit preferences in cost-effectiveness analysis; respondents are asked to make hypothetical tradeoffs. The biggest drawback of this approach is that hypothetical tradeoffs cannot be validated.

4. Discussion

Cost-benefit analysis measures health status in monetary terms, so cost of an intervention can be compared directly to health improvement from an intervention. This is also true for tradeoffs between different components of health outcomes. The monetary values applied to these components are supposed to represent the willingness to pay for each. If so, they can be used to compute rates at which consumers are willing to trade among these components, and to trade each outcome component for other things.

In contrast to cost-effectiveness analyses, some cost-benefit analyses use a revealed preference approach to determining the weights to apply to the various outcomes. That is, weights are inferred from what is revealed about consumers' willingness to pay through actual behavior, rather than through answers to questions about hypothetical choices. In principal, cost-effectiveness studies could use a revealed preference approach as well. Lack of data for situations in which a revealed preference approach could be applied is likely a main reason that such studies rely on hypothetical scenarios, instead. Another is that when real choices are made concerning factors that affect health outcomes, it is often unclear whose preferences the choices reflect – the consumer's, their family's, their employer's, their provider's, their insurer's, a government bureaucrat's, or a politician's.

D. Critique of Conventional Approaches

A common feature of approaches used to measure multi-attribute health outcomes in the cost-effectiveness and cost-benefit literatures is a search for weights by which to value the multiple attributes and aggregate them into a univariate outcome measure. The methods used recognize that consumers are willing to trade among attributes (e.g., better health for a limited period, at the cost of a shortened life), but differ in their approaches to measuring willingness to trade.

As stated earlier, methods for determining weights that rely on hypothetical situations are problematic for the simple reason that individuals' choices when the alternatives are real may be quite different. Hence, we prefer revealed preference approaches, at least conceptually. Many applications are problematic, however, because the information that is used to reveal preferences might not do that. There are several possible reasons:

- Variability in willingness to pay both across individuals and with changes in the individual's circumstances. By and large, both the hypothetical and revealed preference approaches take as given that there is a "correct" set of weights for valuing attributes of health outcomes. Yet there are many reasons to think that consumer or social willingness

to trade varies because of variation in "tastes" and variation in individual circumstances. For example, we know that some individuals are more averse to risk than others, so some would likely be less willing to take a chance on a procedure that will restore health, but with a small probability of immediate death, than others. As another example, consider the value of an individual's time, which varies with the individual's circumstances. Morbidity for a corporate executive may mean significant loss of income for the corporation, as well as for the executive's family, so the value the executive places on ending morbidity may be substantially higher than the value assigned by a person whose activities generate substantially less income. Further, a person's willingness to trade may change over time, as an individual's economic and social circumstances change;

- It is often not clear whose preferences are reflected in actual choices that are made, especially in health care. Choices actually made are often influenced by the preferences of the provider, the insurer, public policy makers, government bureaucrats, employers, and family members, as well as those of the consumer. Revealed preference approaches generally assume that consumers make well-informed choices and are not influenced by the preferences of others;

- Conventional methods do not consider trade-offs in production – the ability of the health system to increase one outcome attribute at the expense of another (e.g., a year of life for three years of improved quality of life). While it seems natural to value outputs on the basis of consumer or societal preferences, much can be said about the efficiency of producing multi-attribute health outcomes without considering preferences at all. If one system produces better outcomes in all dimensions than another, holding resources constant, then the first system is technically more efficient. Of course this example is too simple because, holding resources constant, one system may produce more of one outcome attribute than a comparison system, but less of another. In this situation we can still make meaningful comparisons if we know what the trade-offs between attributes are on the production side. Then we could answer the question: can the first system substitute among outcome attributes to a point where all attributes are at least as great as for the second system. If so, how large would the surplus be? More generally, we can answer questions about efficiency in the production of health outcomes with knowledge of preferences. While such answers beg the question of whether health outcomes are allocated efficiently or equitably in the population, they are important answers. Improving the efficiency with which health outcomes are produced can contribute substantially towards achieving better health outcomes for all; and

- While conventional methods do not consider trade-offs in production explicitly, those based on revealed preference may do so inadvertently, as illustrated by the following example. A consumer might, for instance, be willing to give up a year of life in return for three "high quality years," but it may be that treatment can produce four high quality years per year of life lost. In this situation we might see consumers "paying" for four high quality years with one year of life, but they would be willing to pay the same amount for just three years; i.e., they are willing to pay more. If we use the observed trade-off and count a year of life as equivalent to four years of improved quality, the weights used reflect production trade-offs and not consumer preferences.

III. Hedonic Price Functions

A. Theory

The "hedonic" model was specifically developed to better understand markets for heterogeneous goods – i.e., for goods with varying quantities of one or more utility bearing attributes (Rosen, 1974). In this model, consumers value goods only because of their attributes. Consumers have heterogeneous tastes and incomes, and, given price, choose the good with the attributes that they most prefer. Symmetrically, producers with heterogeneous technologies and/or input prices choose to produce goods with the attributes that, given price, maximize their profits. Under competition, prices adjust to clear the market for goods with each specific set of attributes, and price and cost are equivalent. If the heterogeneous characteristics of consumers and producers have continuous distributions, a continuous multi-dimensional locus, relating price to the good's attributes, emerges – the hedonic price function. The distribution of consumer tastes and producer costs determine where the consumers and producers will choose to locate along the hedonic price function. Therefore, market-clearing prices depend on the distribution of consumer and producer characteristics.

Under the competitive conditions assumed in the theoretical model, the partial derivatives of the hedonic price function with respect to the attributes have an interesting interpretation. The values of the partial derivatives vary along the function; i.e., they depend on the specified attributes of the good. The partial derivative with respect to an attribute represents both the marginal willingness of consumers who purchase goods with the specified attributes to pay for that attribute, and the marginal cost of producing more of that attribute for producers who produce the specified attributes. Because of this interpretation, the partial derivatives are called the "implicit prices" of the attributes. The ratio of the implicit price for one attribute relative to that for another at a specified point is both the rate at which consumers who purchase the good with the specified attributes are willing to substitute one attribute for the other (marginal rate of substitution), and the rate at which the producers are willing to produce one attribute instead of the other (marginal rate of transformation).

Some features of this model are appealing for application to analysis of health outcomes, while others are not. One of the most appealing aspects of the model is the explicit recognition of heterogeneous, multi-attribute outcomes. Also appealing is the idea that both consumers and producers are heterogeneous. For the application of interest to the project, one might specify that the cost of providing health is a function of multi-attribute outcomes such as longevity, morbidity, and quality of life.[1] Consumer characteristics that might affect health outcomes might also be included in the function as cost shifters. Given the theoretical underpinnings of the model, the partial derivatives of the function with respect to the attributes would be interpreted as the implicit prices of the attributes – reflecting both consumer willingness to pay for attributes, at the margin, and provider marginal cost.

[1] The RFP suggests "satisfaction with care" as an attribute. We do not include it, however, because it is more specific attributes of care that generate satisfaction and have value. It would be more appropriate to include attributes that generate satisfaction (e.g. short waiting times).

The problem with this interpretation is that the competitive assumptions underlying the interpretation are not likely to be satisfied in most health applications. Because the health care market is not competitive, the implicit prices from a hedonic price function will not reflect the market's marginal valuation of the attributes.

It would be a heroic, and likely impossible, task to develop a hedonic price model based on assumptions that are reasonable for the health care market. Nonetheless, it is reasonable to posit that observed relationships between cost and attributes of health outcomes reflect an equilibrium that is determined by the interactions of patients, their families, physicians, other providers, hospital owners, third-party payers, regulators, etc. Such an equilibrium relationship would embody the available technologies, the preferences of various actors, and the availability of resources to those actors.

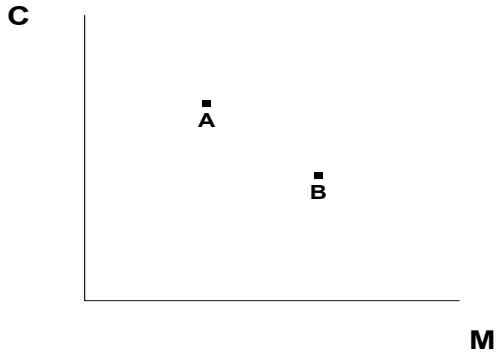
While we should not think of the partial derivatives of the function with respect to the attributes as implicit prices, we can still think of them as the implicit values that are implied by the joint behavior of all participants in the health system. It might well be that some of these implicit values differ markedly from what consumers are willing to pay for the attributes, at the margin, or from the marginal cost of producing the attribute. One might find, for instance, that the implicit values derived from the price function are way out of line with any reasonable notion of consumer willingness to pay, or with marginal cost. Either finding might reasonably be interpreted as an indicator of an inefficient system.

Hedonic price functions from different economic systems could be used as a basis for comparing system performance. If the hedonic price schedule is lower in one system than another over all attribute combinations, then the first system's performance is better than the second's in the sense that those in the first system have to give up less of other things to obtain the same good. In the textbook hedonic model, this situation would reflect differences in technologies. In health care such differences would reflect both differences in technologies and differences in the efficiency of their use.

B. An Illustration

To illustrate the nature of the challenge in measuring the burden of disease, we consider birth outcomes in two different systems. This illustration will be developed empirically in the remainder of the report. For the sake of simplicity, we will consider only two outcomes, infant morbidity and health care cost. Later, we will also consider maternal morbidity. Suppose mean health care cost is higher in system A relative to system B, but that mean infant morbidity is lower (***Exhibit 1***). Which system exhibits better performance at reducing the burden of disease?

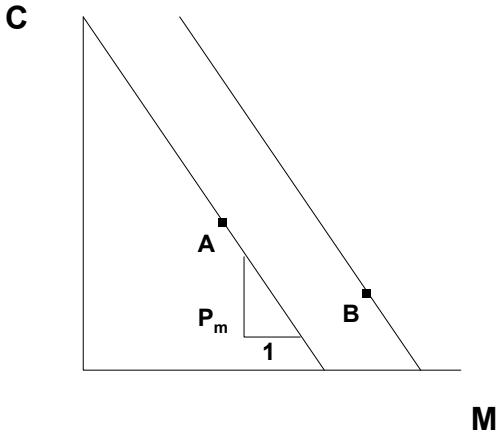
Exhibit 1
Illustration of Outcomes for Two Health Systems



To answer this question, we would first want to know whether patient risk factors are similar for the two populations. Suppose that outcomes have been adjusted for risk in an appropriate fashion (see ***Section V***). As each system exhibits better mean outcomes than the other in one outcome component (infant morbidity or health care cost), a natural next step is to develop a scheme to weight these outcomes, and compare the weighted outcomes. Equivalently, as implied by the hedonic price model, we could develop a scheme to place a dollar value, or implicit price, on infant morbidity and then simply add the dollar value of infant morbidity to the cost of health care (see ***Section IV***) infant morbidity.

If a fixed dollar price P is applied to a unit of infant morbidity, all outcomes that impose the same burden of disease (health care cost plus price times infant morbidity) as those for system A lie along a line that passes through A and has slope minus P (***Exhibit 2***). Outcomes below this line are associated with a lower burden of disease and outcomes above this line are associated with a higher burden of disease. The outcome for system B can lie above or below the system A line depending on the price P associated with infant morbidity.

Exhibit 2
Illustration of Health System Outcome Comparison with a Constant Price for Morbidity

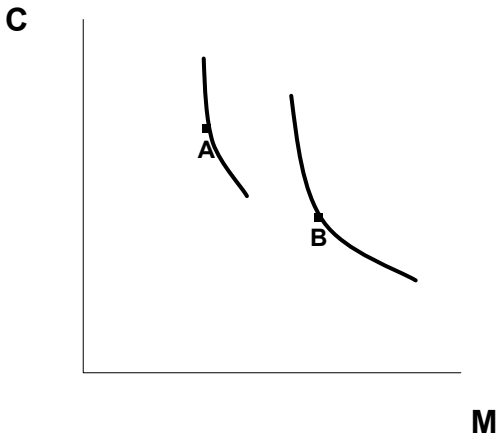


The notion behind the hedonic price approach is that the price a consumer must pay for a good will vary positively with the quantity of any desirable attribute, holding other attributes

constant. In a competitive marketplace, the observed relationship between the price of a good and the good's attributes reflects both the marginal cost of producing the attributes and the willingness of the marginal consumer to pay for the attributes. In a less than fully competitive market, market power and preferences of third parties might play a significant role in determining the relationship between price and the attributes of a good.

If we extend this idea to the birth example, we might find that, within each of the two systems, low morbidity is associated with higher cost, reflecting the notion that reducing morbidity is costly (***Exhibit 3***). We might conclude that the burden of disease is lower in System A than in System B because any (risk-adjusted) level of morbidity can be attained at a lower (risk-adjusted) cost. The higher level of average morbidity in System B reflects the higher cost of attaining any given level of morbidity under System B and a choice not to spend the resources that would be required to attain the average morbidity level observed under System A. If we believe that the observed relationships reflect the true cost of reducing morbidity in each system, it is not necessary to assess individual or social preferences to conclude that System A is more successful at reducing the burden of disease than System B. (See ***Section VI*** for a comparison of two systems).

Exhibit 3
Illustrative Morbidity/Cost Trade-offs for Two Health Systems



The discussion above assumes that there is a negative relationship between morbidity and cost, in the tradition of the hedonic price literature, which assumes that increasing desired attributes is costly, and consumers are willing to pay for such increases. There are, however, two reasons why the relationship between morbidity and cost might be positive, after controlling for risk. The first of these is a phenomenon we call "outcome induced expenditures." In the case of births, even if considerable resources are used to reduce the risk of significant child morbidity, the birth outcome for the child might be poor, and additional treatment resources required (see ***Appendix A***). If these resources were captured in the cost variable, high morbidity would be associated with high cost. The second reason is that some aggressive interventions (e.g., unnecessary C-sections) cause poorer outcomes, and physician knowledge and/or incentives could be such that they such interventions are used even when they are counterproductive. While the first reason is consistent with efficient use of resources, the second is not. A third reason is that the method used to control for risk might be inadequate. As riskier patients have

higher morbidity and higher health care cost than less risky patients, inadequate control for risk could result in a positive relationship morbidity and health care cost that are partially-adjusted for risk, even if fully-adjusted measures would have a negative relationship.

As will be shown below, we do observe a positive relationship between cost and morbidity for risk-adjusted birth outcomes. In such situations, can the relationship between cost and outcome attributes still provide a meaningful yardstick for system performance? We think the answer is yes. In general, if System A can produce any (partially risk-adjusted) outcome at lower cost than System B, it is using fewer resources to obtain the same result. The exception would be if an incomplete risk-adjustment methodology biased the adjusted outcome measures for one system, relative to the other, and in the direction that would invalidate this interpretation.

IV. Estimation

A. Regression Analysis

The economics literature does not have a clear solution for the estimation of hedonic price models. Following Rosen's (1974) lead, estimation typically takes the form of a multiple regression of price on attributes. More recent analysts have recognized that there is an endogeneity problem with this approach (Epple, 1987). Prices and attributes are all jointly determined in the transaction between consumer and producer. Hence, any stochastic factors that affect prices will also likely affect attributes. This simultaneity problem may lead to bias in regression estimates of the implicit prices. The only known solution to the problem that is fully satisfactory from a conceptual perspective is specification and estimation of a full "structural" model, but this is not practical. For practical reasons, analysts have continued to use multiple regression to estimate hedonic price models, despite its shortcomings.

B. Structural Linear Relations

While we are not able to specify and estimate a full structural model, we are dissatisfied with the literature's use of linear regression as the next best alternative because of its asymmetric treatment of the multiple endogenous variables. Hence, we produce alternative estimates using a methodology in which the cost and the morbidity variables are treated symmetrically. Like regression, the method we use is a variant of a structural linear equation method.

Let C represent cost, let the vector $\underline{Y}$ represent a set of endogenous outcomes, and let the vector $\underline{X}$ represent a set of factors that are exogenous to the outcomes. The structural linear model is specified as follows:

Equation 1: $C_i^* = \alpha + \underline{\beta}'\underline{Y}_i^* + \underline{\gamma}'\underline{X}_i$

Equation 2: $C_i = C_i^* + u_i$

Equation 3: $\underline{Y}_i = \underline{Y}_i^* + \underline{v}_i$

where:

- C_i and C_i^* are measured (observed) and actual (unobserved) cost, respectively;

- $\underline{Y}_i$ and $\underline{Y}_i$* are measured (observed) and actual (unobserved) outcomes, respectively;
- u_i and $\underline{v}_i$ are measurement errors for C_i and $\underline{Y}_i$, respectively;
- α is the intercept, to be estimated;
- β is the coefficient vector for $\underline{Y}$, to be estimated; and
- γ is the coefficient vector for $\underline{X}$, to be estimated.

Further, we assume that:

- u_i, and the elements of $\underline{v}_i$ are orthogonal to one another and to C_i*, $\underline{Y}_i$*, and $\underline{X}_i$; and
- u_i, and $\underline{v}_i$ are independently and identically and joint normally distributed across observations.

This model treats C and Y symmetrically, with one arbitrary exception: we have normalized the coefficient of C* in Equation 1 to be unity and put C* and $\underline{Y}$* on opposite sides of the equality. Because of this normalization, we can interpret each element of β as the marginal savings (if negative) or cost (if positive) associated with a marginal unit of the corresponding element of Y.

The three-equation model can easily be reduced to a two-equation model, in which one of the equations is almost identical to the regression model. Substitution of Equation 1 into Equation 2 yields:

Equation 4: $C_i = \alpha + \underline{\beta}'\underline{Y}_i^* + \underline{\gamma}'\underline{X}_i + u_i,$

which, along with the relevant assumptions, is the usual multiple regression model apart from the fact that $\underline{Y}_i$* is not directly observed.

As specified, the model is not identified. That is, without further specification it would not be possible to estimate β. One way to achieve identification is to assume that the variance of v_i, the measurement error for $\underline{Y}_i$, is zero. The model then reduces to the classical regression model, and regression of C on $\underline{Y}$ produces asymptotically unbiased and efficient estimates of α and β. This assumption, however, is asymmetric with respect to the treatment of C and $\underline{Y}$.

Our alternative estimator is based on a model obtained via an identifying assumption that treats C and $\underline{Y}$ symmetrically: the proportion of variation in C_i that is due to u_i is assumed to be identical to the proportion of variation in $\underline{Y}_i$ that is due to v_i.[2]

We used the SAS statistical package to the regression and structural linear relationship analysis. The program used to obtain the results presented in this report is included in

[2] An alternative way to state this assumption is that the error variances for the standardized values of C and Y are identical.

Appendix B. Regressions were estimated using the REG procedure and structural linear relationships were estimated using the CALIS procedure.

V. Data

A. The Birth Product Line Data

The data set used in this study is collected as part of The National Quality Management Program (NQMP), which is the quality improvement and utilization management oversight program of TRICARE. As part of the quality management function, NQMP conducts special studies chosen to represent high volume and high visibility clinical issues. The Birth Product Line (BPL) data set was collected to conduct studies on obstetrics. For this report, we use data collected in 1996.

The Birth Product Line data set has very rich information on both mothers and babies. The data are collected starting from the first prenatal visit until discharge after delivery. All procedures and outcomes within this time period are recorded. Also included in the data set are demographic characteristics of the mother, risk factors of the mother, and hospital characteristics. Resource utilization by the mother and the infant are also recorded.

We consider four key "outcome" variables: morbidity score, resource value units, maternal functional status, and maternal satisfaction.

Morbidity score summarizes the mother's and infant's morbidity following delivery. There are separate indicators for the mother and infant. This variable ranges from zero to 12, with zero representing no morbidity and 12 representing fatality. This variable is derived using physicians' assessment of the morbidity of the patient. For the analysis, we treat the morbidity score variables as if they are continuous, cardinal variables.

Resource value units is the amount that was expended for the mother's and infant's care until discharge after delivery. There are separate accounts for the mother and the infant. The mother's resource value units are recorded between the first prenatal visit and discharge, whereas the infant's resource value units are recorded between birth and discharge.

Functional status summarizes the functional status of the mother after hospital discharge following delivery. Functional status is derived from a questionnaire mailed to the mother after delivery. The scale is based on the Duke Activity Score index and ranges between zero and one, with zero representing bad functional status and one representing good functional status.

Satisfaction score summarizes the satisfaction of the mother with the service received. Satisfaction score is based on a questionnaire mailed to the mother after delivery. The scale is based on the Brigham and Women's Patient Satisfaction Questionnaire and ranges between zero and four, with zero representing excellent service and four representing poor service.

As discussed further, below, we ultimately did not use the last two outcome measures in the analysis because of data limitations.

B. Sample Construction

1. Exclusion of Multiple and Abroad Births

We have excluded multiple births from the analysis (58 observations). The data set contains aggregate infant morbidity and infant cost measures for multiple births, and does not contain any information about how many births there were in a multiple birth, making it impossible to know the per infant morbidity and cost.

We have also excluded abroad births from the analysis (631 observations). These births are excluded because overseas military hospitals predominantly employ local physicians trained outside the United States who might have different decision-making guidelines from physicians trained in the United States.

The distributions of maternal and child morbidity scores after exclusions are almost identical to those before exclusions, as are the means of the maternal and infant cost variables, conditional on morbidity scores (***Exhibit 4***).

Exhibit 4
Comparison of Full Sample to Sample with Exclusions

	Maternal				Infant			
	All Births		Excluding Multiple and Abroad Births		All Births		Excluding Multiple and Abroad Births	
Sample size	**7,701**		**7,021**		**7,701**		**7,021**	
Morbidity	**(%)**	**Mean Log Cost**	**(%)**	**Mean Log Cost**	**(%)**	**Mean Log Cost**	**(%)**	**Mean Log Cost**
0	48.3	5.35	48.3	5.34	48.5	3.61	48.6	3.60
1	0.7	5.42	0.8	5.41	5.2	3.81	5.1	3.81
2	39.2	5.30	39.1	5.29	23.9	3.85	23.8	3.84
3	9.7	5.46	9.7	5.45	7.8	3.96	8.1	3.95
4+	2.1	5.68	2.1	5.67	14.6	4.45	14.4	4.41
Total	100	5.35	100	5.34	100	3.93	100	3.92

Source: Lewin analysis of the Birth Product Line Data.

All findings in the remainder of this report are based on the sample of 7,021 cases remaining after exclusion of multiple and abroad births, or subsets of those observations.

2. Exclusion of Functional Status and Satisfaction Score from the Analysis

As mentioned above, we have excluded functional status and satisfaction score variables from the analysis, because of data limitations. Both variables were obtained from the patients via a mailed assessment form mailed out after delivery. Only 1,339 of the 7,704 patients mailed the survey back (24 percent). The low response rates create two problems: a smaller than desired sample size for the analysis, and the distinct possibility that the sample for which all data are available poorly represents the population.

We found that mothers who respond to the questionnaire tended to have higher morbidity scores and higher mean expenditures, given morbidity scores, than those in the full sample (***Exhibit 5***). We also found that the coefficient of the maternal morbidity in a regression of cost on maternal morbidity changes substantially when the cases with missing functional status and satisfaction score are dropped from the analysis.[3] This indicates that dropping these observations would result in coefficient estimates that are substantially biased. Hence, we do not include models with functional status and satisfaction scores in the final report.

Exhibit 5
Exclusion of Cases with Missing Values for Functional Status and Maternal Satisfaction

	Observations with Missing Functional Status and Satisfaction Score:			
	Included		Excluded	
Sample Size	7,021		1,339	
Maternal Morbidity	**Frequency (%)**	**Mean Log Cost**	**Frequency (%)**	**Mean Log Cost**
0	48.3	5.35	40.1	5.46
1	0.7	5.42	0.6	5.85
2	39.2	5.30	45.3	5.36
3	9.7	5.46	11.7	5.49
4	2.1	5.68	2.3	5.55
Total	100	5.35	100	5.43

Source: Lewin analysis of the Birth Product Line Data.

C. Patient and Hospital Characteristics

1. Patient Risk Factors

The BPL data include actual and risk-adjusted values for morbidity, cost, functional status, and satisfaction score, all in logarithms. The risk adjustment adjusts the outcome and resource use variables for maternal risk factors, including health status and demographic characteristics, as well as a few indicators of the health status of the fetus. For example, a mother with gestational diabetes would require more care, and as a result more resources, than a mother without gestational diabetes.

The regression analyses presented use risk-adjusted residuals of the outcome variables. We constructed "residuals" by subtracting the logarithmic predicted score, which appears in the data set, from the unadjusted logarithmic score.[4] The variation in the risk-adjusted score is the part of the variation in the actual variable that the risk adjusters could not account for.

[3] These regressions used risk-adjusted cost and maternal morbidity variables. Using a sample in which we only excluded observations with multiple or abroad coefficients, the maternal morbidity coefficient was 0.235, with a t-statistic of 20.5. When the observations with missing functional status and satisfaction scores are dropped, the coefficient drops to 0.149 and the t-statistic to 5.9.

[4] See ***Appendix C*** for further discussion of the risk adjustment process.

2. Hospital Characteristics

As part of our analysis, we test for differences between observed trade-offs across hospitals. In order to determine if the observed differences can be explained by hospital characteristics, in some of the analyses we use morbidity and cost variables that have been adjusted for hospital characteristics. To obtain the adjusted values, we regressed each variable on the hospital characteristics, and then computed the residual.[5] The residual contains the variation in the analysis variables that the hospital characteristics cannot account for.

VI. Analysis and Results

A. Model Specification

We explore the observed relationships between health care costs and birth outcomes using the BPL data. We initially use the approach that is commonly applied in the econometric literature on hedonic price functions: multiple regressions of cost on infant and maternal morbidity. Because costs and morbidity are jointly determined, however, we also use structural linear equations to treat cost and infant and maternal morbidity symmetrically. In both cases, the variables have all been risk-adjusted for maternal characteristics. We also adjusted the variables for hospital characteristics in some instances, as indicated.

We use the risk adjusted cost and morbidity variables in both the regressions and structural linear relationship models. The regression models use adjusted cost as the dependent variable and adjusted maternal morbidity and adjusted infant morbidity as the independent variables.

Separate models are estimated for each of three cost measures: total, maternal and infant. Because we view the maternal and infant outcomes of the birth to be determined jointly, we focus on the total cost equation. The separate equations for maternal and infant costs are of interest, however, because all infant costs are incurred after delivery, and from that point on treatment for the infant is essentially independent of treatment for the mother. Nonetheless, there might be a relationship between maternal morbidity and infant cost, after risk adjustment, because treatments that adversely affect maternal morbidity (e.g., a Cesarean section) might be used to improve infant morbidity, and also reduce the cost of the infant's post-partum care. Such a relationship would likely disappear if infant morbidity and cost were risk-adjusted for the infant's status at birth, but the risk-adjusted measures we use are adjusted for pre-delivery risk factors only.

The regression models are specified as follows:

Equation 5: $C_M = \beta_{MM} M_M + \beta_{MI} M_I + u_M$

Equation 6: $C_I = \beta_{IM} M_M + \beta_{II} M_I + u_I$

[5] See ***Appendix C*** for further discussion of the risk adjustment process.

Equation 7: $C_T = \beta_{TM} M_M + \beta_{TI} M_I + u_T$

where C_M, C_I, and C_T are the log of maternal, infant and total cost, after risk adjustment;

M_M and M_I are risk-adjusted maternal and infant morbidity.

Other assumptions follow those described in Section IV.B.

The regression model treats cost and morbidity score asymmetrically, assuming causality goes in the direction of morbidity score to health care cost. It does not take into account the joint determination of cost and morbidity. To reflect the joint determination of the outcome variables, we use a structural linear relationship model to treat all outcome variables symmetrically, as endogenous variables determined within the system by the exogenous variables.

To set up the structural linear relationship model, we assume that morbidity score is measured with random error, following the framework developed in ***Section IV***. This assumption is justified by the fact that the process through which the morbidity measures are determined is subjective. Initially, a physician (or a team of physicians) assigns a morbidity score to every condition a patient has. Morbidity scores assigned to a patient are later aggregated. Both the score assignment and aggregation processes are subjective.

The assumptions for the structural linear model follow those described in ***Section IV***. The error terms are assumed to be independent of all true values, and of one another. The total cost model alone has three equations, one for each endogenous variable:[6]

$$M_M = M_M^* + u_M$$

$$M_I = M_I^* + u_I$$

$$C_T = \beta_{TM} M_M^* + \beta_{TI} M_I^* + u_T$$

In addition, we specify that the variance of each error is proportional to the variance of the corresponding measured variable.[7] The models for maternal and child cost are symmetric.

B. Entire Sample

In the total cost equation, the coefficients on both morbidity variables are positive and quite significant (***Exhibit 8***). Because all variables are in logs, the coefficients can be interpreted as

[6] At first glance the model appears to be asymmetric with respect to the three outcome variables, but this is because we have arbitrarily normalized the coefficient of cost to be unity. Results would be identical after if we were change the normalization and appropriately transform the parameters.

[7] To implement this restriction, we standardize all of the variables and restrict the error variances of the three standardized variables to equal one another. The coefficients reported are rescaled to reflect the unstandardized variables.

elasticities.[8] Thus, a one percent increase in maternal morbidity is associated with a 0.2 percent increase in cost, and a one percent increase in infant morbidity is associated with a 0.4 percent increase in cost.

Both morbidity coefficients are also positive in each of the maternal and infant cost equations. The coefficient of maternal morbidity is the higher of the two in the maternal cost equation, and the coefficient of infant morbidity is higher in the infant equation.

Regressions for the entire sample have an intercept of zero, because both the dependent and independent variables are residuals with mean zero. Because the error term also has a mean of zero, it follows that the intercept has to be zero.

Exhibit 8
Results from Regression Analysis

	Dependent Variable		
Independent Variables	**Maternal Cost**	**Infant Cost**	**Total Cost**
Constant	0	0	0
Maternal Morbidity (risk adjusted)	0.235 (20.5)	0.118 (13.1)	0.188 (18.1)
Infant Morbidity (risk adjusted)	0.137 (11.9)	0.642 (71.5)	0.447 (43.0)
R^2	0.079	0.438	0.249
Sample Size	7021	7021	7021

Source: Lewin analysis of the Birth Product Line Data.

The morbidity coefficients are uniformly higher in the structural linear relationships (SLR) than in the regression estimates (***Exhibit 9***). The maternal and infant morbidity elasticities in the total cost equation are 0.4 and 0.9 respectively, more than twice as large as in the regressions. The higher coefficients in the structural linear relationships reflect the symmetric identifying assumption on the error variances. That is, if we assume that the share of variation in the morbidity indices that is noise is as large as the share of variation in cost that is noise (after risk adjustment), the strength of the relationship between the underlying "true" variables appears stronger, than if we assume that all of the variation in the morbidity indices is real, and the only unexplained variation is in the cost variable. This is a common result in errors-in-variables models. While both of these identifying assumptions are probably incorrect, it seems reasonable to conclude that (a) the partial relationships between both morbidity scores and cost are positive, holding the other score constant; and (b) they are stronger than the regression coefficients suggest, because the latter are based on the extreme assumption of no error variation in the morbidity variables. The SLR coefficient of infant morbidity in the infant cost equation is exceptionally large – an elasticity just under 1.0.

The t-statistics for the SLR coefficients are very close to the corresponding values for the regression coefficients. This is because both models are just identified. The R^2 values from the

[8] Elasticities measure the percent increase in one variable associated with a one percent increase in another.

SLRs are higher than for the regression estimates, because they are estimates of the variation in the unobserved true cost variable explained by the unobserved true morbidity variables. This is another reflection of the fact that the identifying assumption of this model makes the relationships between the true variables appear stronger than the identifying assumption for the regression model.

Exhibit 9
Results from Structural Linear Relations Analysis

	Dependent Variable		
Independent Variables	**Maternal Cost**	**Infant Cost**	**Total Cost**
Constant	0	0	0
Maternal Morbidity (risk adjusted)	0.856 (19.5)	0.179 (13.1)	0.383 (17.8)
Infant Morbidity (risk adjusted)	0.499 (11.2)	0.975 (71.6)	0.909 (42.5)
R^2	0.271	0.654	0.485
Sample Size	7,021	7,021	7,021

Source: Lewin analysis of the Birth Product Line Data.
Note: R^2 for this model is the share of the variation in the unobserved true cost variable that is accounted for by the unobserved true morbidity variables.

All of these estimates clearly show a positive relationship between risk adjusted morbidity and cost. As discussed previously, this could be because of: (a) outcome induced expenditures; (b) aggressive interventions that cause poorer outcomes; and/or (c) inadequate risk adjustment. Outcome induced expenditures are especially likely to be the cause of the high infant morbidity coefficient because even after all reasonable care is taken, given pre-natal risk, some infants will have poor outcomes and substantial resources will be used to address them. This is especially evident in the infant cost equation, which reflects post-natal expenses for the infant only, because infant morbidity is not adjusted for the infant's status at birth.

We explore the reason for the positive morbidity coefficients further in the next section.

C. Differences In Relationship Across Delivery Types

If the positive relationship between cost and morbidity is a result of outcome induced expenditures or of inadequate risk adjustment, then the relationship between cost and morbidity should differ across delivery types; because the types of patients who have different types of delivery should have different degrees of risk factors or adverse outcomes. For example, one explicit source of outcome-induced expenditures is an emergency C-section. That is, the physician anticipates a vaginal delivery, something goes wrong, and an emergency measure is taken to respond to an intermediate outcome that might be closely related to the final outcome. This suggests that if we use the vaginal-delivery sub sample only, the problem of outcome-induced expenditures should be reduced, at least for maternal outcomes – we would have eliminated cases with an important outcome that can induce significant expenditures. Put differently, among vaginal deliveries we would expect less variation in maternal morbidity and less variation in costs associated with variation in morbidity. This might also be true among mothers who had a planned C-section (cases that have a planned "bad" intermediate outcome

for the mother), and even among mothers with an emergency C-section, but this is less clear. It is more difficult to predict the effect on infant morbidity coefficients.

The regressions using sub-samples do not have an intercept of zero, because the dependent and independent variables do not have a mean of zero for the sub-sample, only for the entire sample.

For total cost, the maternal morbidity coefficient is lower in the vaginal delivery equation than in the full sample, but higher in both the elective and emergency C-section samples (***Exhibits 10*** and ***11***). This result is true for both the regression and SLR estimates. The smaller value in the vaginal delivery sample does suggest that one reason for the positive coefficient in the full sample is outcome-induced expenditures, but the coefficient remains positive.

The relationship between the infant morbidity coefficients in the sub-sample and full-sample estimates depends on the estimation method. For emergency C-sections, the coefficient is smaller than for the full sample, regardless of estimation method.

Exhibit 10
Results from Regression Analysis by Delivery Type: Total Cost

	Sample			
Independent Variables	**Full Sample**	**Vaginal Delivery**	**Elective C-Section**	**Emergency C-Section**
Constant	0	-0.084 (-6.9)	0.211 (4.8)	0.421 (12.2)
Maternal Morbidity (risk adjusted)	0.188 (18.1)	0.110 (8.2)	0.320 (5.5)	0.255 (9.6)
Infant Morbidity (risk adjusted)	0.447 (43.0)	0.435 (33.5)	0.534 (13.8)	0.387 (16.4)
R^2	0.249	0.196	0.378	0.336
Sample Size	7,021	4,947	357	833

Source: Lewin analysis of the Birth Product Line Data.

Exhibit 11
Results from Structural Linear Relations Analysis by Delivery Type: Total Cost

	Sample			
Independent Variables	**Full Sample**	**Vaginal Delivery**	**Elective C-Section**	**Emergency C-Section**
Constant	0	-0.084 (-7.0)	0.211 (5.6)	0.421 (14.4)
Maternal Morbidity (risk adjusted)	0.383 (17.8)	0.267 (8.2)	0.723 (6.1)	0.373 (11.3)
Infant Morbidity (risk adjusted)	0.909 (42.5)	1.005 (33.2)	0.770 (13.8)	0.575 (16.9)
R^2	0.485	0.455	0.655	0.567
Sample Size	7,021	4,947	357	833

Source: Lewin analysis of the Birth Product Line Data.
Note: R^2 for this model is the share of the variation in the unobserved true cost variable that is accounted for by the unobserved true morbidity variables.

The relation of sub-sample coefficients to full-sample coefficients for both morbidity variables in the maternal cost equations depends on estimation method (***Exhibits 12*** and ***13***). It is difficult to detect patterns that lend themselves to interpretation about the causes of the positive morbidity coefficients.

Exhibit 12
Results from Regression Analysis by Delivery Type: Maternal Cost

	Sample			
Independent Variables	**Full Sample**	**Vaginal Delivery**	**Elective C-Section**	**Emergency C-Section**
Constant	0	-0.097 (-7.1)	0.173 (3.6)	0.488 (13.5)
Maternal Morbidity (risk adjusted)	0.235 (20.5)	0.145 (9.8)	0.381 (5.9)	0.325 (11.7)
Infant Morbidity (risk adjusted)	0.137 (11.9)	0.106 (7.4)	0.267 (6.2)	0.131 (5.3)
R^2	0.079	0.030	0.166	0.184
Sample Size	7,021	4,947	357	833

Source: Lewin analysis of the Birth Product Line Data.

Exhibit 13
Results from Structural Linear Relations Analysis by Delivery Type: Maternal Cost

	Sample			
Independent Variables	**Full Sample**	**Vaginal Delivery**	**Elective C-Section**	**Emergency C-Section**
Constant	0	-0.097 (-7.2)	0.173 (4.1)	0.489 (16.0)
Maternal Morbidity (risk adjusted)	0.856 (19.5)	1.086 (8.8)	1.290 (6.6)	0.525 (13.8)
Infant Morbidity (risk adjusted)	0.499 (11.2)	0.642 (6.3)	0.435 (6.2)	0.216 (5.5)
R^2	0.271	0.217	0.518	0.470
Sample Size	7,021	4,947	357	833

Source: Lewin analysis of the Birth Product Line Data.
Note: R^2 for this model is the share of the variation in the unobserved true cost variable that is accounted for by the unobserved true morbidity variables.

It is also difficult to see patterns in comparisons of morbidity coefficients in the infant cost equations (***Exhibits 14*** and ***15***). One finding, however, is that the maternal morbidity coefficient in the infant cost equation is smaller in the vaginal delivery sample than in the full delivery sample, regardless of estimation method.

Exhibit 14
Results from Regression Analysis by Delivery Type: Infant Cost

	Sample			
Independent Variables	**Full Sample**	**Vaginal Delivery**	**Elective C-Section**	**Emergency C-Section**
Constant	0	-0.065 (-6.2)	0.270 (7.2)	0.421 (12.2)
Maternal Morbidity (risk adjusted)	0.118 (13.1)	0.058 (5.0)	0.187 (3.8)	0.255 (9.6)
Infant Morbidity (risk adjusted)	0.642 (71.5)	0.652 (58.6)	0.618 (18.6)	0.387 (16.4)
R^2	0.438	0.413	0.501	0.336
Sample Size	7,021	4,947	357	833

Source: Lewin analysis of the Birth Product Line Data.

Exhibit 15
Results from Structural Linear Relations Analysis by Delivery Type: Infant Cost

	Sample			
Independent Variables	**Full Sample**	**Vaginal Delivery**	**Elective C-Section**	**Emergency C-Section**
Constant	0	-0.065 (-6.3)	0.270 (8.4)	0.296 (10.9)
Maternal Morbidity (risk adjusted)	0.179 (13.1)	0.093 (5.1)	0.330 (4.3)	0.213 (7.7)
Infant Morbidity (risk adjusted)	0.975 (71.6)	1.025 (58.7)	0.811 (19.1)	0.739 (26.0)
R^2	0.654	0.651	0.719	0.657
Sample Size	7,021	4,947	357	833

Source: Lewin analysis of the Birth Product Line Data.
Note: R^2 for this model is the share of the variation in the unobserved true cost variable that is accounted for by the unobserved true morbidity variables.

Overall, we find it difficult to draw any strong conclusion about the reason for the positive morbidity coefficients from the analysis of the sub-samples defined by delivery type. Perhaps the most important finding is that for each sub-sample the coefficients on both morbidity variables are positive and quite significant in all equations, regardless of estimation method – another indication of the strength of the positive empirical relationship between risk-adjusted morbidity and risk-adjusted cost in the BPL data.

VII. Differences Between Hospital Types

A. Why Trade-offs May Be Different Across Different Hospital Types

The relationship between cost and morbidity may differ across hospitals due to patient mix, management, incentives, facility, equipment, physician preferences for morbidity, physician skill and knowledge, and other reasons. Risk adjustment should, at least partially, correct for patient mix, but not other factors.

We do not have enough data to estimate models for each hospital, but we do have enough data to test for differences in relationships by hospital type. We compare civilian to military hospitals and, among military hospitals, major medical centers, large community hospitals, and small community hospitals.

In each case we first test whether just the coefficients of the morbidity variables are constant across hospitals of different types. This test allows for the possibility of differing intercepts, but constant slopes (i.e., parallel planes for cost and maternal/infant morbidity relationships), after risk adjustment. We also test for equality of the slopes and the intercepts.

B. Relationship between Cost and Morbidity in Military and Civilian Hospitals

Military hospitals have higher health care cost by morbidity score than civilian hospitals (***Exhibit 16***). This could mean that military hospitals are less efficient or that the patients at military hospitals are at higher risk.

Exhibit 16
Descriptive Statistics for Civilian and Military Hospitals

	Civilian Mother Morbidity		Military Mother Morbidity		Civilian Infant Morbidity		Military Infant Morbidity	
Score	**Obs (%)**	**Average Log Cost**	**Obs (%)**	**Average Log Cost**	**Obs (%)**	**Average Log Cost**	**Obs (%)**	**Average Log Cost**
0	60.1	5.14	43.6	5.45	51.3	3.39	47.5	3.70
1	1.4	5.21	0.5	5.63	4.2	3.49	5.5	3.91
2	29.6	5.06	42.9	5.36	23.5	3.65	23.9	3.91
3	7.3	5.19	10.6	5.52	6.4	3.84	8.7	3.98
4	1.6	5.49	2.4	5.70	14.6	4.37	14.4	4.42
Total	100	5.13	100	5.43	100	3.77	100	3.98

Source: Lewin analysis of the Birth Product Line Data.

Coefficients on maternal morbidity are higher for civilian hospitals than for military hospitals, but differences are not large except in the total cost equation (***Exhibits 17*** through ***19***). Coefficients on infant morbidity are lower for military hospitals than for civilian hospitals in some models, but higher in others; again differences are not large.

Exhibit 17
Estimated Relationships for Civilian and Military Hospitals: Maternal Cost

	Regressions		Structural Linear Relations	
Independent Variables	**Civilian Hospitals**	**Military Hospitals**	**Civilian Hospitals**	**Military Hospitals**
Constant	-0.456 (-23.4)	0.187 (14.2)	-0.456 (23.6)	0.187 (14.2)
Maternal Morbidity (risk adjusted)	0.284 (13.6)	0.193 (14.9)	0.850 (13.5)	0.626 (14.3)
Infant Morbidity (risk adjusted)	0.143 (7.7)	0.133 (9.9)	0.304 (7.6)	0.548 (9.4)
R^2	0.117	0.065	0.445	0.242
Sample Size	2007	5014	2007	5014

Source: Lewin analysis of the Birth Product Line Data.

Exhibit 18
Estimated Relationships for Civilian and Military Hospitals: Infant Cost

	Regressions		Structural Linear Relations	
Independent Variables	**Civilian Hospitals**	**Military Hospitals**	**Civilian Hospitals**	**Military Hospitals**
Constant	-0.167 (-9.8)	0.070 (6.8)	-0.167 (-9.9)	0.070 (6.8)
Maternal Morbidity (risk adjusted)	0.158 (8.7)	0.094 (9.4)	0.242 (8.73)	0.141 (9.4)
Infant Morbidity (risk adjusted)	0.776 (48.1)	0.580 (55.1)	1.073 (48.2)	0.904 (55.2)
R^2	0.554	0.391	0.763	0.602
Sample Size	2007	5014	2007	5014

Source: Lewin analysis of the Birth Product Line Data.

Exhibit 19
Estimated Relationships for Civilian and Military Hospitals: Total Cost

	Regressions		Structural Linear Relations	
Independent Variables	**Civilian Hospitals**	**Military Hospitals**	**Civilian Hospitals**	**Military Hospitals**
Constant	-0.329 (-17.4)	0.136 (11.5)	-0.329 (-17.5)	0.136 (11.5)
Maternal Morbidity (risk adjusted)	0.233 (11.5)	0.154 (13.4)	0.933 (31.2)	0.299 (13.3)
Infant Morbidity (risk adjusted)	0.567 (31.6)	0.391 (32.4)	0.459 (11.4)	0.828 (32.2)
R^2	0.375	0.206	0.645	0.427
Sample Size	2007	5014	2007	5014

Source: Lewin analysis of the Birth Product Line Data.

C. Testing for Differences in the Relationship Between Cost and Morbidity Between Civilian and Military Hospitals

We performed F-tests for two sets of restrictions on the military and civilian hospital regressions (***Exhibit 20***). For the first, we tested that the coefficients of the maternal and infant morbidity scores ("slopes") are the same in the two types of hospitals. This set of restrictions could not be rejected. For the second, we tested whether both the slopes and intercepts were the same. This set of restrictions was clearly rejected.

Exhibit 20
F-Tests for Slope and Intercept Restrictions: Civilian and Military Hospitals

	R-Squared for Restrictions Indicated			F-Statistic	
Dependent Variable	**None**	**Slope**	**Slope and Intercept**	**Slope**	**Slope and Intercept**
Maternal Cost	0.1664	0.1648	0.0791	44.7	257.1***
Infant Cost	0.4594	0.4498	0.4379	87.3	97.7**
Total Cost	0.3020	0.2937	0.2485	59.3	191.1***

Source: Lewin analysis of the Birth Product Line Data.
Note: ** Reject at 95%. *** Reject at 99%.

Given the acceptance of the slope restrictions, this analysis provides an example of how these models can be used to compare the efficiency of two hospital types. When just the slope restrictions are imposed (***Exhibit 21***), the exponentiated difference between the military and civilian intercepts minus one is an approximate measure of the percent difference in costs, holding risk-adjusted maternal and child morbidity constant. The regression results imply that total costs at military hospitals are 60 percent higher than at civilian hospitals.[9]

Exhibit 21
Regressions with Restricted Slopes and Unrestricted Intercepts for Civilian and Military Hospitals

Independent Variables	Maternal Cost	Infant Cost	Total Cost
Constant	-0.464 (-22.7)	-0.173 (-10.5)	-0.337 (-17.9)
Maternal Morbidity (risk adjusted)	0.215 (19.6)	0.110 (12.4)	0.174 (17.2)
Infant Morbidity (risk adjusted)	0.137 (12.5)	0.642 (72.2)	0.447 (44.4)
Military Hospital	0.649 (26.8)	0.243 (12.4)	0.471 (21.2)
R^2	0.165	0.450	0.294
Sample Size	7,021	7,021	7,021

Source: Lewin analysis of the Birth Product Line Data.

[9] This figure was calculated as exp(.471) –1.

D. Relationship Between Cost and Morbidity in Medical Centers, Large Community Hospitals, and Small Community Hospitals

Military hospitals can be further divided into three groups: medical centers (MCs), large community hospitals (LCHs), and small community hospitals (SCHs). The BPL data include sufficient observations to estimate separate cost models for each type, and to test for equality of the slopes and intercepts.

SCHs had more low-risk mothers than the others, but also had a substantial number of high risk mothers (***Exhibit 22***). Maternal risk distributions are about the same for LCHs and MCs. Risk differences are much greater for infants. While 54 percent of SCH infants were low risk, only 50 percent of LCH infants and 41 percent of MC infants were low risk (***Exhibit 23***).

In all three hospital types, mean costs increase with both maternal and infant morbidity. LCHs and MCs have somewhat higher mean maternal costs than SCHs within maternal morbidity categories, but differences are not large. Within infant morbidity categories, SCHs and LCHs have comparable infant costs, and in most categories their costs are *higher* than MC costs. While mean infant costs overall are higher at MCs than at either SCHs or LCHs, this reflects a combination of higher morbidity infants and lower costs within risk categories.

Exhibit 22
Descriptive Statistics: Mother Morbidity by Military Hospital Type

	SCH		LCH		MC	
	Obs (%)	**Average Log Cost**	**Obs (%)**	**Average Log Cost**	**Obs (%)**	**Average Log Cost**
0	47.2	5.41	42.5	5.46	42.6	5.48
1	0.2	5.42	0.2	5.82	1.1	5.61
2	42.4	5.31	44.2	5.37	41.8	5.38
3	8.4	5.50	10.9	5.53	11.7	5.51
4	1.8	5.63	2.2	5.82	2.8	5.60
Total	100	5.38	100	5.44	100	5.45

Source: Lewin analysis of the Birth Product Line Data.

Exhibit 23
Descriptive Statistics: Infant Morbidity by Military Hospital Type

	SCH		LCH		MC	
Infant Score	**Obs (%)**	**Average Log Cost**	**Obs (%)**	**Average Log Cost**	**Obs (%)**	**Average Log Cost**
0	54.4	3.70	49.5	3.72	40.7	3.66
1	5.7	4.01	5.7	3.93	5.1	3.80
2	25.1	3.92	24.3	3.93	22.8	3.88
3	5.9	4.09	7.2	4.08	12.2	3.88
4	8.9	4.57	13.3	4.54	19.2	4.20
Total	100	3.88	100	3.97	100	4.07

Source: Lewin analysis of the Birth Product Line Data.

As in earlier models, we continue to find positive relationships between morbidity scores and cost, after risk adjustment (***Exhibits 24*** through ***26***). With one exception, we find that the coefficients of the morbidity score variables are quite similar across the three types of hospitals. The exception is the coefficient of infant morbidity in the infant cost equation, which is much higher in the MCs than in the LCHs or the SCHs, in both the regression and SLR models. This difference in the infant cost equation is reflected to some extent in the total cost equation.

Exhibit 24
Estimated Relationships by Military Hospital Type: Maternal Cost

	Regressions			Structural Linear Relations		
Independent Variables	**MC**	**LCH**	**SCH**	**MC**	**LCH**	**SCH**
Constant	0.238 (10.2)	0.193 (9.6)	0.085 (3.3)	0.238 (10.3)	0.193 (9.6)	0.085 (3.3)
Maternal Morbidity (risk adjusted)	0.204 (9.4)	0.165 (8.6)	0.223 (7.9)	0.652 (9.1)	0.502 (8.2)	0.869 (7.6)
Infant Morbidity (risk adjusted)	0.155 (7.0)	0.125 (6.0)	0.072 (2.4)	0.502 (6.8)	0.607 (5.7)	0.387 (2.3)
R^2	0.077	0.052	0.059	0.274	0.213	0.235
Sample Size	1,773	2,104	1,137	1,773	2,104	1,137

Source: Lewin analysis of the Birth Product Line Data.

Exhibit 25
Estimated Relationships by Military Hospital Type: Infant Cost

	Regressions			Structural Linear Relations		
Independent Variables	**MC**	**LCH**	**SCH**	**MC**	**LCH**	**SCH**
Constant	0.063 (3.4)	0.074 (4.9)	0.018 (0.9)	0.063 (3.4)	0.074 (4.9)	0.018 (0.9)
Maternal Morbidity (risk adjusted)	0.103 (6.0)	0.082 (5.6)	0.104 (4.9)	0.146 (6.0)	0.121 (5.6)	0.170 (4.9)
Infant Morbidity (risk adjusted)	0.731 (42.1)	0.496 (31.0)	0.409 (18.3)	1.038 (42.5)	0.799 (42.5)	0.707 (18.5)
R^2	0.304	0.153	0.117	0.725	0.521	0.423
Sample Size	1,773	2,104	1,137	1,773	2,104	1,137

Source: Lewin analysis of the Birth Product Line Data.

Exhibit 26
Estimated Relationships by Military Hospital Type: Total Cost

	Regressions			Structural Linear Relations		
Independent Variables	**MC**	**LCH**	**SCH**	**MC**	**LCH**	**SCH**
Constant	0.176 (8.4)	0.132 (7.5)	0.035 (1.5)	0.176 (8.5)	0.132 (7.5)	0.035 (1.5)
Maternal Morbidity (risk adjusted)	0.163 (8.3)	0.134 (7.9)	0.177 (7.1)	0.292 (8.3)	0.257 (7.8)	0.407 (7.0)
Infant Morbidity (risk adjusted)	0.510 (25.8)	0.321 (17.3)	0.243 (9.3)	0.920 (25.9)	0.739 (17.2)	0.640 (9.2)
R^2	0.304	0.153	0.117	0.550	0.344	0.276
Sample Size	1,773	2,104	1,137	1,773	2,104	1,137

Source: Lewin analysis of the Birth Product Line Data.

E. Testing for Differences in the Relationship Between Cost and Morbidity Between Medical Centers, Large Community Hospitals, and Small Community Hospitals

When we tested the equality of the slope coefficients, we found this restriction to be clearly rejected in the infant cost equation, marginally rejected in the total cost equation, and accepted in the maternal cost equation – not surprising giving the pattern of coefficients on the infant morbidity variable noted above (***Exhibit 27***).

Exhibit 27
F-tests for Slope and Intercept Restrictions for Military Hospital Types

	R-Squared for Restrictions Indicated			F-Statistic	
Dependent Variable	**None**	**Slope**	**Slope and Intercept**	**Slope**	**Slope and Intercept**
Maternal Cost	0.1047	0.1011	0.0791	10.0	35.6***
Infant Cost	0.4641	0.4418	0.4379	101.4***	59.6***
Total Cost	0.2775	0.2591	0.2485	61.3**	51.8***

Source: Lewin analysis of the Birth Product Line Data.
Note: **Reject at 95% ***Reject at 99%

While the rejection of the slope restrictions makes it problematic to use the model with restricted slopes and unrestricted intercepts to compare the efficiency of these three hospital types, we proceed to do so for illustrative purposes. Conditional on the slope restrictions, the estimates show that total costs are higher at both SCHs and LCHs than at MCs, after controlling for risk and maternal and infant morbidity – by 16 percent at SCHs and 26 percent at LCHs (***Exhibit 28***).

Exhibit 28
Regressions with Restricted Slopes, Unrestricted Intercepts by Military Hospital Type

Independent Variables	Maternal Cost	Infant Cost	Total Cost
Constant	-0.134 (-8.7)	-0.057 (-4.7)	-0.090 (-6.6)
Maternal Morbidity (risk adjusted)	0.231 (20.4)	0.116 (13.0)	0.186 (18.0)
Infant Morbidity (risk adjusted)	0.145 (12.8)	0.645 (72.0)	0.453 (43.8)
Small Community Hospital	0.228 (7.1)	0.103 (4.1)	0.152 (5.2)
Large Community Hospital	0.325 (12.6)	0.136 (6.7)	0.228 (9.7)
R^2	0.101	0.442	0.259
Sample Size	7,021	7,021	7,021

Source: Lewin analysis of the Birth Product Line Data.
Note: The F-test rejected the slope restrictions in the infant and total cost equations.

F. Relationship Between Cost and Morbidity in Medical Centers, Large Community Hospitals, and Small Community Hospitals after Adjusting for Hospital Characteristics

Some of the variation in the cost/morbidity relationships across the three hospital types could be due to differences in specific observable characteristics of the hospitals. To assess this possibility, we adjusted the cost and morbidity measures for observed hospital characteristics as well as maternal risk factors, and re-estimated the models with the revised cost and

morbidity variables.[10] We found only very small changes in the slope coefficients (compare ***Exhibits 29*** through ***31*** to ***Exhibits 24*** through ***26***). Tests of both the slope and intercept restrictions show that they continue to be strongly rejected by the revised data (***Exhibit 32***).

Exhibit 29
Estimated Relationships by Military Hospital Type after Adjusting for Hospital Characteristics: Maternal Cost

	Regressions			Structural Linear Relations		
Independent Variables	**MC**	**LCH**	**SCH**	**MC**	**LCH**	**SCH**
Constant	0.029 (1.2)	0.020 (1.0)	-0.082 (-3.0)	0.029 (1.2)	0.020 (1.0)	-0.082 (-3.0)
Maternal Morbidity (risk adjusted)	0.213 (9.3)	0.182 (8.9)	0.236 (7.7)	0.797 (8.8)	0.680 (8.3)	1.229 (7.4)
Infant Morbidity (risk adjusted)	0.175 (8.0)	0.132 (6.2)	0.060 (1.9)	0.539 (7.5)	0.606 (5.8)	0.355 (1.8)
R^2	0.083	0.056	0.055	0.272	0.213	0.283
Sample Size	1,773	2,104	1,137	1,773	2,104	1,137

Source: Lewin analysis of the Birth Product Line Data.

Exhibit 30
Estimated Relationships by Military Hospital Type after Adjusting for Hospital Characteristics: Infant Cost

	Regressions			Structural Linear Relations		
Independent Variables	**MC**	**LCH**	**SCH**	**MC**	**LCH**	**SCH**
Constant	0.018 (0.9)	0.013 (0.8)	-0.059 (-2.9)	0.018 (0.9)	0.013 (0.8)	-0.059 (-2.9)
Maternal Morbidity (risk adjusted)	0.112 (5.9)	0.090 (5.6)	0.117 (5.1)	0.168 (5.9)	0.142 (5.6)	0.210 (5.1)
Infant Morbidity (risk adjusted)	0.778 (42.9)	0.525 (31.7)	0.419 (18.0)	1.123 (42.9)	0.864 (31.7)	0.768 (18.0)
R^2	0.520	0.336	0.246	0.744	0.544	0.436
Sample Size	1,773	2,104	1,137	1,773	2,104	1,137

Source: Lewin analysis of the Birth Product Line Data.

[10] See ***Appendix C*** for further discussion of the risk adjustment process.

Exhibit 31
Estimated Relationships by Military Hospital Type after Adjusting for Hospital Characteristics: Total Cost

	Regressions			Structural Linear Relations		
Independent Variables	**MC**	**LCH**	**SCH**	**MC**	**LCH**	**SCH**
Constant	0.025 (1.1)	0.016 (0.9)	-0.073 (-3.0)	0.025 (1.1)	0.016 (0.9)	-0.073 (-3.0)
Maternal Morbidity (risk adjusted)	0.174 (8.2)	0.149 (8.0)	0.193 (7.1)	0.346 (8.1)	0.331 (7.9)	0.546 (6.8)
Infant Morbidity (risk adjusted)	0.552 (26.9)	0.338 (17.5)	0.241 (8.8)	1.017 (26.7)	0.820 (17.3)	0.724 (26.7)
R-squared	0.318	0.157	0.110	0.575	0.360	0.303
Sample Size	1,773	2,104	1,137	1,773	2,104	1,137

Source: Lewin analysis of the Birth Product Line Data.

Exhibit 32
F-tests for Slope and Intercept Restrictions after Controlling for Hospital Characteristics

	R-Squared for Restrictions Indicated			F-Statistic	
Dependent Variable	**None**	**Slope**	**Slope and Intercept**	**Slope**	**Slope and Intercept**
Maternal Cost	0.0685	0.0665	0.0644	50.0**	5.4
Infant Cost	0.4164	0.3950	0.3940	89.2**	46.7***
Total Cost	0.2233	0.2075	0.2059	49.4**	27.2***

Source: Lewin analysis of the Birth Product Line Data.
Note: ** Reject at 95% *** Reject at 99%

VIII. Conclusion

In this paper we have examined one of the main challenges in measuring the burden of disease: the fact that the burden of disease has multiple components (medical costs, reduced quality of life, reduced longevity, productivity loss, etc.). We have reviewed relevant literature, discussed various empirical approaches to measuring trade-offs between the various components of the burden of disease within health systems, and considered the problem of comparing performance across health systems when the outcome has multiple components. While the economics literature has successfully applied the hedonic price function approach to markets for many multi-component products, the theory that supports that approach assumes a level of competition that is not found in the health care industry. Analysis of multi-component outcomes is further complicated in the health care industry because of risk and because outcomes that are worse than expected can result in costs that are higher than expected, to cover treatment for the adverse outcome.

Many empirical approaches can be used to estimate trade-offs between various outcome components, but none are well supported by theory. Use of regression to estimate cost models,

with outcome components as explanatory variables, is inappropriate because outcomes and costs are jointly determined. In principle, fully identified structural models should be applied, but the instrumental variables needed to estimate such models are often not available.

To illustrate the theoretical and statistical issues, we explored the trade-off between risk-adjusted morbidity and risk-adjusted cost in deliveries, using the Birth Product Line data. We used two statistical approaches to illustrate the importance of treating outcome components and costs symmetrical. Following most of the hedonic price literature, we first estimate cost regressions that treat cost and outcomes asymmetrically. Cost is the dependent variable, and the only source of stochastic error in the model is in the determination of cost. Then we estimate structural linear relationships (SLRs) in which independent stochastic errors affect all outcomes and cost. To identify the model, we assume that the variance of each stochastic error is proportional to the variance of the observed variable. This assumption treats outcomes and costs symmetrically, but is not well founded in theory and might be no more defensible than the assumption for the regression model.

What we find is that the relationship between morbidity and cost, after risk adjustment, is positive; cost increases as morbidity increases. This relationship is consistent across delivery types (vaginal, elective c-section, emergency c-section), hospital systems (military, civilian), and hospital types (small community, large community, medical center). We find the same qualitative result in both the regression and SLR estimates. The relationships are stronger in the SLR estimates, reflecting the assumption that risk-adjusted morbidity, as well as cost, is affected by unobserved stochastic factors.

There are at least three possible reasons for the positive relationship between morbidity and cost. One reason is that high costs partially reflect attempts to mitigate bad outcomes from risky events. We refer to this phenomenon as outcome induced expenditures. The scenarios under which outcome induced expenditures can arise are discussed in detail in ***Appendix A***. A second reason is that some aggressive interventions (e.g., unnecessary C-sections) cause poorer outcomes, and physician knowledge and/or incentives are such that they might use such interventions even when they are counterproductive. A third reason is inadequate risk adjustment in the data. Riskier patients have higher morbidity and higher health care cost than less risky patients. Inadequately controlling for risk leads to a spurious correlation between morbidity and health care cost.

Whatever the reasons for the positive relationship between morbidity and costs, it is interesting to compare the nature of these relationships across hospitals of different types. When we compare civilian and military hospitals, we cannot reject the hypothesis that the risk-adjusted trade-offs between morbidity and costs are the same, but we can reject that they are at the same level. That is, holding risk constant, a given increase in morbidity is associated with about the same increase in cost at the two types of hospitals, but costs at military hospitals are proportionately higher at all morbidity levels – by about 60 percent.

When we compare types of military hospitals (medical centers, large community hospitals, and small community hospitals), we find similar tradeoffs between maternal morbidity and cost, after risk adjustment, but infant morbidity is associated with larger increases in cost at medical centers than at the other two types of hospitals. Despite this difference, however, it appears that

cost per case at medical centers is lower than at the two types of community hospitals, holding risk-adjusted morbidity constant.

To sum up, it seems critical to recognize that the observed relationship between the multiple components of healthcare outcomes, including cost, are a product of the interplay between production and consumer preference factors when considering the burden of disease. The entire relationship is of relevance to when comparing the effectiveness of health systems, not just individual outcomes. Yet our ability to understand how this relationship is determined, in theory, and to appropriately estimate it in practice is limited. Because costs are jointly determined with other outcomes, estimation of cost regressions is not the right approach. Approaches that treat cost and other outcomes symmetrically yield different answers, but there are many such approaches and no clear criteria to select among them.

The problem can be simplified by assuming that observed outcomes reflect only production. We did not take this approach here, because it ignores the role that consumers play in making health care decisions. We discuss methods to estimate production relationships between cost and morbidity in detail in ***Appendix D***.

Selected Bibliography

Drummond, Michael F. et al. *Methods for the Economic Evaluation of Health Care Programmes.* Oxford: Oxford University Press, 1997.

Eisenberg, John M. "Clinical Economics: A Guide to the Economic Analysis of Clinical Practices." Journal of the American Medical Association 262 (1989): 2879-2886.

Epple, Dennis. "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products." Journal of Political Economy 95 (1987): 59-80.

Folland, Sherman, Allen C. Goodman, and Miron Stano. *The Economics of Health and Health Care*. 2d ed. Upper Saddle River: Prentice-Hall, 1997.

Gold, Marthe R. et al. "Identifying and Valuing Outcomes." In *Cost-Effectiveness in Health and Medicine*. Oxford: Oxford University Press, 1996.

Goodman, Clifford. "Introduction to Health Care Technology Assessment: Fundamental Concepts and Issues." http://www.nlm.nih.gov/nichsr/ta101/ta10104.htm#Heading16.

Grosskopf, Shawna and Vivian G. Valdmanis. "Measuring Hospital Performance: A Non-Parametric Approach." Journal of Health Economics 6 (1987): 89-107.

Guralnik, Jack M., Linda P. Fried, and Marcel E. Salive. "Disability as a Public Health Outcome in the Aging Population." Annual Review of Public Health 17 (1996): 25-46.

Jones, Andrew M. "Health Econometrics." In *Handbook of Health Economics*. Vol. 1. North Holland: Elsevier Science, forthcoming.

Hodgson, Thomas A. "Cost of Illness in Cost-Effectiveness Analysis: A Review of the Methodology." PharmacoEconomics 6 (1994): 536-52.

Kooreman, Peter. "Data Envelopment Analysis and Parametric Frontier Estimation: Complementary Tools." Journal of Health Economics 13 (1994): 345-46.

Maddala, G. S. *Limited-Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs, No. 3. Cambridge: Cambridge University Press, 1997.

Murray, Christopher J. L. "Rethinking DALYs." In *The Global Burden of Disease.* Global Burden of Disease and Injury Series, vol. 1. Cambridge: Harvard University Press, 1996.

Newhouse, Joseph P. "Frontier Estimation: How Useful A Tool for Health Economics?" Journal of Health Economics 13 (1994): 317-22.

Patrick, Donald L. and Marilyn Bergner. "Measurement of Health Status in the 1990s." Annual Review of Public Health 11 (1990): 165-83.

Rosen, Sherwin. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." Journal of Political Economy 82 (1974): 34-55.

Skinner, Jonathan. "What Do Stochastic Frontier Cost Functions Tell Us About Inefficiency?" Journal of Health Economics 13 (1994): 323-28.

Sloan, Frank A., ed. *Valuing Health Care*: *Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies*. Cambridge: Cambridge University Press, 1996.

Stoto, Michael A. "Public Health Assessment in the 1990s." Annual Review of Public Health 13 (1992): 59-78.

Valdmanis, Vivian G. "Ownership and Technical Efficiency in Hospitals." Medical Care 28 (1990): 552-561.

Varian, Hal R. *Microeconomic Analysis.* 3d ed. New York: Norton & Company, 1992.

Vitaliano, Donald F. and Mark Toren. "Frontier Analysis: A Reply to Skinner, Dor and Newhouse." Journal of Health Economics 13 (1994): 341-43.

Zuckerman, Stephen, Jack Hadley, and Lisa Iezzoni. "Measuring Hospital Efficiency with Frontier Cost Functions." Journal of Health Economics 13 (1994): 255-80.

Appendix A: Trade-offs Between Cost and Morbidity

In this appendix we present a simple model for births to illustrate the theoretical relationship between cost and morbidity. Specifically, we consider the role that risk plays in determining the observed relationship between cost and morbidity. Risk means that for each treatment decision there is a known distribution of possible outcomes. We show that, in the absence of risk, we should only observe a negative relationship between cost and morbidity; that is, preference variation could result in some decision makers accepting higher morbidity in exchange for lower costs. Our empirical findings indicate the opposite, however. When risk is introduced, the observed relationship between morbidity and cost can be positive. This is true even when the patient population is homogeneous. Heterogeneity in risks within the patient population increases the likelihood that a positive relationship between morbidity and costs will be observed.

Suppose there is only one risk factor that can affect morbidity. In what follows, we will make assumptions about the morbidity distribution for this risk factor, with and without treatment. In a homogeneous population, we will assume that everyone has this risk factor; in a heterogeneous population, we will assume that only a part of the population has this risk factor.

The decision making process is a sequential one. There are three periods: pre-admission (when the mother has outpatient visits), pre-delivery in hospital (when the mother is admitted to the hospital, but before deliver), and post-delivery in hospital (after the mother has delivered the infant and through hospital discharge). In each period, a decision is made to treat or not treat the risk factor. We assume the risk factor can be treated in each period. The effect of the treatment is the difference between the morbidity distributions with and without treatment for the risk factor.

For current purposes, we can simplify the model by combining the first two pre-delivery periods. We also do not need to distinguish between infant and maternal morbidity. We end up with two periods, one before and one after delivery, in which treatment decisions are made, costs are incurred, and outcomes are realized. Let C, T, and M represent cost, treatment and morbidity variables, and index them by 1 and 2 for pre- and post-delivery periods. Let R represent the risk factor.

We can express the relationships between the six variables using the following equations:

$C_1 = c_1(T_1)$,

$E(M_1| T_1, R) = m_1(T_1 | R)$

$M_1 = E(M_1| T_1, R) + \varepsilon$

$C_2 = c_2(T_2)$, and

$E(M_2) = m_2(T_2, M_1| R)$,

where E() is the expected value operator, $m_i()$ and $c_i()$ are morbidity and cost functions for the two periods (i = 1, 2), and e is the random component of M_1, after conditioning on T_1 and R.

We now demonstrate how these four equations generate an observed relationship between cost and morbidity under various scenarios.

Scenario 1: Homogeneity with no risk. We start with a homogeneous population where everyone has the risk factor. We make the following morbidity distribution assumptions.

$\Pr (M_1 = 0 \mid T_1 = 1) = 1$

$\Pr (M_1 = 2 \mid T_1 = 0) = 1$

$\Pr (M_2 = 1 \mid T_2 = 1, M_1 = 2) = 1$

If treated before delivery, there is no morbidity at delivery; if not there is severe morbidity. In the latter case, treatment after delivery results in moderate morbidity at discharge. If we assume that low morbidity and low costs are desired outcomes, then costs determine treatment decisions under this scenario. If pre-natal treatment is less expensive than post-natal treatment, then pre-natal treatment will be given; we should only observe low-cost outcomes with no morbidity. If post-natal treatment is less expensive, then pre-natal treatment will depend on the relative value that the decision-makers place on no morbidity and moderate morbidity. Outcomes will either be no morbidity paired with high cost, or modest morbidity paired with low cost. Preference variation can only result in a negative relationship between cost and morbidity, contrary to what we observe.

The above paragraph assumes the risk factor is perfectly treatable, i.e. pre-natal treatment leads to no morbidity. This might seldom be the case. Suppose the following assumptions apply, instead:

$\Pr (M_1 = 1 \mid T_1 = 1) = 1,$

$\Pr (M_1 = 3 \mid T_1 = 0) = 1,$

$\Pr (M_2 = 0 \mid T_2 = 1, M_1 = 1) = 1,$

$\Pr (M_2 = 2 \mid T_2 = 1, M_1 = 3) = 1,$

Even if treated before delivery, there is still moderate morbidity, and therefore some need for treatment after delivery. Here again the decision to treat in each period depends on treatment costs and the relative value of morbidity.

Under homogeneity with no risk, from the decision-maker's perspective the morbidity of each patient is a deterministic function of T_1 and T_2. T_1 and T_2 in turn are determined by treatment costs and the valuation of morbidity. So, it is clear that if lower morbidity is cheaper, treatment will be undertaken. It is only if lower morbidity is more expensive that there will be a decision to be made. Preference variation might result in an observed negative relationship between morbidity and costs; i.e., more health care is associated with better outcomes. Our observation that higher morbidity is associated with higher costs is not explained by this scenario.

Scenario 2: Homogeneity with risk. Now we introduce risk to the above model. Risk means that for each treatment decision there is a known distribution of possible outcomes.

We replace the deterministic equations for morbidity with probability distributions. Suppose the following assumptions apply:

$\Pr(M_1 = 0 \mid T_1 = 1) = 0.6$,

$\Pr(M_1 = 2 \mid T_1 = 1) = 0.4$,

$\Pr(M_1 = 0 \mid T_1 = 0) = 0.3$,

$\Pr(M_1 = 2 \mid T_1 = 0) = 0.7$, and

$\Pr(M_2 = 1 \mid T_2 = 1, M_1 = 2) = 1$.

As in scenario 1, the decision to treat will depend on treatment costs and the decision-maker's valuation of morbidity. The difference between scenario 1 and 2 is that morbidity is not deterministic function of T_1 and T_2 anymore.

This scenario can explain the association between high morbidity with high cost. Use of T_1 will again depend on relative costs of T_1 and T_2, and the decision-maker's valuation of morbidity. Consider a provider who always uses T_1. Two types of discharge outcomes will be observed: no morbidity paired with the cost of T_1, alone (Pr = 0.6), and modest morbidity paired with the cost of T_1 and T_2, combined (Pr = 0.4). If, instead the provider never uses T_1, the two discharge outcomes will be no morbidity paired with no cost (Pr = 0.3), and modest morbidity paired with the cost of T_2 alone (Pr = 0.7). For each of these providers, there will be a negative relationship between C and M across all cases.

Whether we observe a positive or negative relationship in data for multiple providers who share the same outcome expectations, but have varying preferences, depends on the cost of T_1, and the effect that T_1 has on reducing the probability of a bad outcome for M_1. If T_1 is very efficacious and high cost relative to T_2, then we are likely to observe a negative relationship between C and M because expected costs for the providers who always use T_1 will be higher than for those who do not. If T_1 is inexpensive, and/or not very efficacious, the opposite will be true.

Provider uncertainty, or lack of knowledge, about risks could also induce variation in outcomes. Variation in knowledge has an effect that is similar to preference variation. If all have the same preferences, but have differing assessments of risk, some will always choose T_1, but others will never choose it.

In sum, once risk is added to the model, a positive association between costs and morbidity can be generated, even if the patient population is homogeneous – but a negative association is also possible.

Scenario 3: Heterogeneity and risk adjustment. Heterogeneity in risk factors within a patient population makes a positive relationship between cost and morbidity more likely, as

those at high risk are likely to receive more treatment and have worse outcomes because of the risk. We rely on risk adjustment to control for the risk factor. Risk adjustment of an outcome (cost or morbidity) means subtraction of the expected difference between the outcome for an individual patient's risk group and the outcome for the population from the outcome for the patient.

For risk adjustment to work well, the average outcome must capture adequately the effects of variation in the risk factor on outcomes. If the risk factor varies within each risk group, it will induce a positive relationship between morbidity and cost within the group, which in turn contributes to a positive relationship between the (partially) risk-adjusted costs and morbidity for all patients in the population.

Appendix B: SAS Programs

SAMPLE.SAS

```
/* This program produces descriptive statistics for the full and estimation samples */
libname bod 'R:\project\Burden of Disease\1897.05';

data bpl96 (drop=pbscore);
  set bod.bpl96;
  civil=(rgn=.); military=(1<=rgn<=12); abroad=(13<=rgn<=15);
  patient=_N_;

  if lbtrvu=. and tbscore=12 then do; lbtrvu=5.82; ltotrvu=6.07; end;
  if lbtrvu=. and tbscore=24 then do; lbtrvu=6.20; ltotrvu=6.43; end;

data ussingle;
  set bpl96;
  if mulbrth=0 and abroad=0;

data civilian military small large medical missing;
  set ussingle;
  if military=1 then output military; if military=0 then output civilian;
  if smlcom=1 then output small;   if lrgcom=1 then output large;
  if medctr=1 then output medical;
  if funsta^=. and satis^=. then output missing;

proc freq data=missing; tables mscore;
proc tabulate data=missing; class mscore; var lmtrvu; table mscore all, lmtrvu*mean;
proc freq data=bpl96; tables mscore tbscore; title 'all';
proc freq data=ussingle; tables mscore tbscore; title 'est';
proc tabulate data=bpl96; class mscore; var lmtrvu; table mscore all, lmtrvu*mean; title 'all';
proc tabulate data=ussingle; class mscore; var lmtrvu; table mscore all, lmtrvu*mean; title 'est';
proc tabulate data=bpl96; class tbscore; var lbtrvu; table tbscore all, lbtrvu*mean; title 'all';
proc tabulate data=ussingle; class tbscore; var lbtrvu; table tbscore all, lbtrvu*mean; title 'est';

proc freq data=military; tables mscore tbscore; title 'mil';
proc freq data=civilian; tables mscore tbscore; title 'civ';
proc tabulate data=military; class mscore; var lmtrvu; table mscore all, lmtrvu*mean; title 'mil';
proc tabulate data=civilian; class mscore; var lmtrvu; table mscore all, lmtrvu*mean; title 'civ';
proc tabulate data=military; class tbscore; var lbtrvu; table tbscore all, lbtrvu*mean; title 'mil';
proc tabulate data=civilian; class tbscore; var lbtrvu; table tbscore all, lbtrvu*mean; title 'civ';
```

ADJUST.SAS

```
/* This program risk-adjusts the cost and morbidity variables using patient characteristics and
estimates the relationship between cost and mornidity using linear regression and structural
linear relationships */

libname bod 'R:\project\Burden of Disease\1897.05';

data bpl96 (drop=pbscore);
  set bod.bpl96;
  civil=(rgn=.); military=(1<=rgn<=12); abroad=(13<=rgn<=15);
  patient=_N_;

  if lbtrvu=. and tbscore=12 then do; lbtrvu=5.82; ltotrvu=6.07; end;
  if lbtrvu=. and tbscore=24 then do; lbtrvu=6.20; ltotrvu=6.43; end;

data ussingle;
  set bpl96;
  if mulbrth=0 and abroad=0;

/* Risk adjusting */
data ussingle;
  set ussingle;

%LET ADJUST =     adhbw adlbw anhbw anmlpr anolhy anplcprv eduhi faageg35
                  faagel18 fafasian fafnavy fafotrsv fafracms fafwhite famact
                  famblack famusaf fapayle4 hsanchor hsancrlg hsanedma hsang42w
                  hsanmlbr hsaspreg hsasthma hsbleed3 hscord hscsect hsdabges
                  hsdepres hsetohb hshpv hshrtotr hshtn hspdhkd hspeclmo hspeclsv
                  hspyelo hsstill hsthyrod hsvagdel opgesl37 opmalprs ouanem5
                  ouedma1 ouedma4 ougbstrp ouglycu3 ouhtn3 oumafpab oupapsmr ourh
                  ousvedm4 phherpad phsvedma pranbplv prprom rdnstabn;

%LET HOSPITAL = brthlo brthmed brthhi oblev1 oblev2 oblev3 hitech broom nonresp
                  mdpbd rnpbd rnrtio;

proc means mean std data=ussingle;
var &ADJUST;

proc reg data=ussingle noprint; model lmtrvu=&ADJUST;
id patient;
output out=mcost predicted=pmcost residual=rmcost;
title 'mother cost';

proc reg data=ussingle noprint; model lbtrvu=&ADJUST;
id patient;
output out=bcost predicted=pbcost residual=rbcost;
```

```
title 'baby cost';

proc reg data=ussingle noprint; model ltotrvu=&ADJUST;
id patient;
output out=tcost predicted=ptcost residual=rtcost;
title 'total cost';

proc reg data=ussingle noprint; model tbscore=&ADJUST;
id patient;
output out=bscore predicted=pbscore residual=rbscore;
title 'baby score';

proc reg data=ussingle noprint; model revmscr4=&ADJUST;
id patient;
output out=mscore predicted=pmscore residual=rmscore;
title 'mother score';

proc reg data=ussingle noprint; model revtscr4=&ADJUST;
id patient;
output out=tscore predicted=ptscore residual=rtscore;
title 'total score';

proc sort data=ussingle; by patient;
proc sort data=mcost; by patient;
proc sort data=bcost; by patient;
proc sort data=tcost; by patient;
proc sort data=mscore; by patient;
proc sort data=bscore; by patient;
proc sort data=tscore; by patient;

data us;
  merge ussingle mcost bcost tcost mscore bscore tscore;
  by patient;

/* Standardizing the residuals */
proc standard data=us std=1 out=usstand;
var rmscore rmcost rbscore rtscore rbcost rtcost;

data military civilian opvagdel opelcsec opemcsec small large medical;
      set usstand;
      if military=1 then output military;
      if military=0 then output civilian;
      if opvagdel=1 then output opvagdel;
      if opelcsec=1 then output opelcsec;
      if opemcsec=1 then output opemcsec;
      if smlcom=1 then output small;
```

```
      if lrgcom=1 then output large;
      if medctr=1 then output medical;

/* Linear Regression Analysis */
%MACRO regress (depvar, dataset);
proc reg data=&dataset; model &depvar=rmscore rbscore; title "&dataset";
%mend;

%regress (rmcost, usstand); %regress (rbcost, usstand); %regress (rtcost, usstand);
%regress (rmcost, military); %regress (rbcost, military); %regress (rtcost, military);
%regress (rmcost, civilian); %regress (rbcost, civilian); %regress (rtcost, civilian);
%regress (rmcost, opvagdel); %regress (rbcost, opvagdel); %regress (rtcost, opvagdel);
%regress (rmcost, opemcsec); %regress (rbcost, opemcsec); %regress (rtcost, opemcsec);
%regress (rmcost, opelcsec); %regress (rbcost, opelcsec); %regress (rtcost, opelcsec);
%regress (rmcost, small); %regress (rbcost, small); %regress (rtcost, small);
%regress (rmcost, large); %regress (rbcost, large); %regress (rtcost, large);
%regress (rmcost, medical); %regress (rbcost, medical); %regress (rtcost, medical);


/* Structural Linear Relationships Analysis */
%MACRO strlin (depvar, dataset);
proc calis ucov aug data=&dataset;
lineqs
      rbscore=f2+d2,
      rmscore=f1+d1,
      &depvar=alpha intercept + beta f1 + gamma f2 + e1;
std e1=the1, d1=the1, d2=the1, f1=the2, f2=the3;
cov e1 d1=0, e1 d2=0, d1 d2=0;
var rbscore rmscore rtscore rbcost rmcost rtcost military smlcom lrgcom;
title "&dataset: &depvar";
%mend;

%strlin (rmcost, usstand); %strlin (rbcost, usstand); %strlin (rtcost, usstand);
%strlin (rmcost, military); %strlin (rbcost, military); %strlin (rtcost, military);
%strlin (rmcost, civilian); %strlin (rbcost, civilian); %strlin (rtcost, civilian);
%strlin (rmcost, opvagdel); %strlin (rbcost, opvagdel); %strlin (rtcost, opvagdel);
%strlin (rmcost, opemcsec); %strlin (rbcost, opemcsec); %strlin (rtcost, opemcsec);
%strlin (rmcost, opelcsec); %strlin (rbcost, opelcsec); %strlin (rtcost, opelcsec);
%strlin (rmcost, small); %strlin (rbcost, small); %strlin (rtcost, small);
%strlin (rmcost, large); %strlin (rbcost, large); %strlin (rtcost, large);
%strlin (rmcost, medical); %strlin (rbcost, medical); %strlin (rtcost, medical);

/* Testing for Differences in the Relationship Between Cost and Morbidity */
/* Between Civilian and Military Hospitals                                */
data usstand;
  set usstand;
```

```
  mmil=military*rmscore; bmil=military*rbscore;
  msml=smlcom*rmscore;   bsml=smlcom*rbscore;
  mlrg=lrgcom*rmscore; blrg=lrgcom*rbscore;

proc reg data=usstand; model rmcost=rmscore rbscore; title 'mother cost';
proc reg data=usstand; model rmcost=rmscore rbscore military; title 'mother cost';
proc reg data=usstand; model rmcost=rmscore rbscore military mmil bmil; title 'mother cost';

proc reg data=usstand; model rbcost=rmscore rbscore; title 'baby cost';
proc reg data=usstand; model rbcost=rmscore rbscore military; title 'baby cost';
proc reg data=usstand; model rbcost=rmscore rbscore military mmil bmil; title 'baby cost';

proc reg data=usstand; model rtcost=rmscore rbscore; title 'total cost';
proc reg data=usstand; model rtcost=rmscore rbscore military; title 'total cost';
proc reg data=usstand; model rtcost=rmscore rbscore military mmil bmil; title 'total cost';

proc reg data=usstand; model rmcost=rmscore rbscore; title 'mother cost';
proc reg data=usstand; model rmcost=rmscore rbscore smlcom lrgcom; title 'mother cost';
proc reg data=usstand; model rmcost=rmscore rbscore smlcom lrgcom msml bsml mlrg blrg;
title 'mother cost';

proc reg data=usstand; model rbcost=rmscore rbscore; title 'baby cost';
proc reg data=usstand; model rbcost=rmscore rbscore smlcom lrgcom; title 'baby cost';
proc reg data=usstand; model rbcost=rmscore rbscore smlcom lrgcom msml bsml mlrg blrg;
title 'baby cost';

proc reg data=usstand; model rtcost=rmscore rbscore; title 'total cost';
proc reg data=usstand; model rtcost=rmscore rbscore smlcom lrgcom; title 'total cost';
proc reg data=usstand; model rtcost=rmscore rbscore smlcom lrgcom msml bsml mlrg blrg;
title 'total cost';

/* Risk adjusting using hospital characteristics */
data ussingle;
  set bpl96;
  if mulbrth=0 and abroad=0 and military=1;

proc reg data=ussingle noprint; model lmtrvu=&ADJUST &HOSPITAL;
id patient;
output out=mcost predicted=pmcost residual=rmcost;
title 'mother cost';

proc reg data=ussingle noprint; model lbtrvu=&ADJUST &HOSPITAL;
id patient;
output out=bcost predicted=pbcost residual=rbcost;
title 'baby cost';
```

```
proc reg data=ussingle noprint; model ltotrvu=&ADJUST &HOSPITAL;
id patient;
output out=tcost predicted=ptcost residual=rtcost;
title 'total cost';

proc reg data=ussingle noprint; model tbscore=&ADJUST &HOSPITAL;
id patient;
output out=bscore predicted=pbscore residual=rbscore;
title 'baby score';

proc reg data=ussingle noprint; model revmscr4=&ADJUST &HOSPITAL;
id patient;
output out=mscore predicted=pmscore residual=rmscore;
title 'mother score';

proc reg data=ussingle noprint; model revtscr4=&ADJUST &HOSPITAL;
id patient;
output out=tscore predicted=ptscore residual=rtscore;
title 'total score';

proc sort data=ussingle; by patient;
proc sort data=mcost; by patient;
proc sort data=bcost; by patient;
proc sort data=tcost; by patient;
proc sort data=mscore; by patient;
proc sort data=bscore; by patient;
proc sort data=tscore; by patient;

data us;
  merge ussingle mcost bcost tcost mscore bscore tscore;
  by patient;

proc standard data=us std=1 out=usstand;
var rmscore rmcost rbscore rtscore rbcost rtcost;

/* Testing for Differences in the Relationship Between Cost and Morbidity */
/* Between Different Types of Military Hospitals                          */
data usstand;
  set usstand;
  msml=smlcom*rmscore;
  bsml=smlcom*rbscore;
  mlrg=lrgcom*rmscore;
  blrg=lrgcom*rbscore;

data small large medical;
  set usstand;
```

```
  if smlcom=1 then output small;
  if lrgcom=1 then output large;
  if medctr=1 then output medical;

proc reg data=usstand; model rmcost=rmscore rbscore; title 'mother cost';
proc reg data=usstand; model rmcost=rmscore rbscore smlcom lrgcom; title 'mother cost';
proc reg data=usstand; model rmcost=rmscore rbscore smlcom lrgcom msml bsml mlrg blrg;
title 'mother cost';

proc reg data=usstand; model rbcost=rmscore rbscore; title 'baby cost';
proc reg data=usstand; model rbcost=rmscore rbscore smlcom lrgcom; title 'baby cost';
proc reg data=usstand; model rbcost=rmscore rbscore smlcom lrgcom msml bsml mlrg blrg;
title 'baby cost';

proc reg data=usstand; model rtcost=rmscore rbscore; title 'total cost';
proc reg data=usstand; model rtcost=rmscore rbscore smlcom lrgcom; title 'total cost';
proc reg data=usstand; model rtcost=rmscore rbscore smlcom lrgcom msml bsml mlrg blrg;
title 'total cost';

/* Linear regression using residuals adjusted by patient and hospital characteristics */
%MACRO regress (depvar, dataset);
proc reg data=&dataset; model &depvar=rmscore rbscore; title “&dataset”;
%mend;

%regress (rmcost, small); %regress (rbcost, small); %regress (rtcost, small);
%regress (rmcost, large); %regress (rbcost, large); %regress (rtcost, large);
%regress (rmcost, medical); %regress (rbcost, medical); %regress (rtcost, medical);

/* Structural Linear Relationships Analysis */
/* using residuals adjusted by patient and hospital characteristics */
%MACRO strlin (depvar, dataset);
proc calis ucov aug data=&dataset;
lineqs
      rbscore=f2+d2,
      rmscore=f1+d1,
      &depvar=alpha intercept + beta f1 + gamma f2 + e1;
std e1=the1, d1=the1, d2=the1, f1=the2, f2=the3;
cov e1 d1=0, e1 d2=0, d1 d2=0;
var rbscore rmscore rtscore rbcost rmcost rtcost military smlcom lrgcom;
title “&dataset: &depvar”;
%mend;

%strlin (rmcost, small); %strlin (rbcost, small); %strlin (rtcost, small);
%strlin (rmcost, large); %strlin (rbcost, large); %strlin (rtcost, large);
%strlin (rmcost, medical); %strlin (rbcost, medical); %strlin (rtcost, medical);
```

Appendix C: Risk Adjustment

A. Adjustment for Patient Characteristics

We were dissatisfied with the risk-adjusted morbidity and cost variables in the BPL data. When we regressed the risk-adjusted variables on the risk adjusters themselves, we found that the latter had some explanatory power and could not reject the hypothesis that at least some of the coefficients were not zero. Some individual coefficients were also statistically significant. We think the reason for this is that a complex, non-linear method was used in the adjustment process. While that procedure might be preferred to a simpler, linear method for some purposes, it is problematic when the adjusted variables are to be used in linear models such as those estimated here, because the adjusted variables are not orthogonal to the risk adjusters.

We therefore decided to derive alternative, linear risk-adjusted variables. We used a linear regression to adjust the log morbidity score and log cost variables using all risk adjusters that are listed in Appendix A-1 of the BPL documentation. ***Exhibit C.1*** presents descriptive statistics for the patient characteristics that we adjusted for. We regressed each of the variables to be adjusted on all of the risk adjusters. The residuals from these regressions are the alternative risk-adjusted variables.

Exhibit C.1

Patient Characteristic	Statistics for Mothers	
	Mean	Standard Deviation
High birth weight (intrapartum)	0.0305	0.17
Low birth weigh (intrapartum)	0.0050	0.07
High birth weigh (antepartum)	0.0393	0.19
Malpresentation (antepartum)	0.0457	0.21
Oligohydramnios (antepartum)	0.0353	0.18
Placenta previa (antepartum)	0.0232	0.15
College and graduate level, either parent	0.1313	0.34
Patient age>35	0.0457	0.21
Patient age<18	0.0216	0.15
Father-Asian	0.0164	0.13
Father-Navy	0.2065	0.40
Father-other services	0.0376	0.19
Father-race missing	0.5253	0.50
Father-white	0.3250	0.47
Patient-active duty	0.1500	0.36
Patient-black	0.1545	0.36
Patient-Air Force	0.1599	0.37
Pay grade less than E5	0.4452	0.50
Chorioamnionitis (antepartum)	0.0053	0.07
cerclage/incompetent (antepartum)	0.0051	0.07
Edema (antepartum)	0.0632	0.24
>42 wks gestation (antepartum)	0.0047	0.07
History-infertility/assisted pregnancy	0.0162	0.13

Patient Characteristic	Statistics for Mothers	
	Mean	**Standard Deviation**
History-asthma	0.0783	0.27
Third trimester bleeding (antepartum)	0.0121	0.11
Pregnancy history-cord complications	0.0048	0.07
Pregnancy history-previous c-section	0.1125	0.32
History-gestational diabetes	0.0501	0.22
History-psych disorder	0.0275	0.16
History-alcohol use before pregnancy	0.0610	0.24
History-condyloma/HPV	0.0393	0.19
History-other heart conditions	0.0383	0.19
History-hypertension	0.0802	0.27
History-half pack per day during pregnancy	0.0370	0.19
History-preeclampsia, mild/NOS	0.0379	0.19
History-preeclampsia, severe	0.0077	0.09
History-pyelonephritis during pregnancy	0.0078	0.09
Pregnancy history-stillbirth	0.0107	0.10
History-thyroid disorder	0.0192	0.14
Pregnancy history-previous vaginal delivery	0.4445	0.50
Gestational age at delivery<37	0.0610	0.24
Malpresentation at time of delivery	0.0771	0.27
Anemia based on Hgb/Hct, last l	0.0348	0.18
Mild/moderate edema, 1st visit	0.0084	0.09
Mild/moderate edema, last visit	0.1206	0.33
Group B strep (+);	0.0598	0.24
Glycosuria, 2nd to last visit	0.0061	0.08
High BP at 2nd to last visit	0.1602	0.37
MSAFP abnormal	0.0234	0.15
Pap smear abnormal	0.1082	0.31
Rh antibody screen (+)/Rh (-)	0.1474	0.35
Severe edema, last visit	0.0128	0.11
Genital herpes on admission	0.0402	0.20
Severe edema on admission	0.0375	0.19
Procedure-abnormal pelvis	0.0225	0.15
Procedure-premature rupture of membrane	0.0244	0.15
NST abnormal	0.0131	0.11

B. Adjustment for Hospital Characteristics

For some of the analyses, we used cost and morbidity variables that had been adjusted for both patient and hospital characteristics. Hospital characteristics might explain some of the difference in the relationship between cost and morbidity between different types of hospitals, after controlling for patient characteristics.

Hospital characteristics were available only for military hospitals. To perform the adjustment, we started with the morbidity and cost variables adjusted for patient characteristics, using the alternative linear procedure described above. These were each regressed on the hospital characteristics. The residuals from these regressions are morbidity and cost adjusted for both patient and hospital characteristics. ***Exhibit C.2*** presents descriptive statistics for the military hospital characteristics.

Exhibit C.2: Military Hospital Characteristics

Hospital Characteristic	Statistics for Births	
	Mean	**Standard Deviation**
Number of Annual Births < 320	0.10	0.30
320 < Number of Annual Births < 1250	0.43	0.49
1250 < Number of Annual Births < 2550	0.24	0.43
Obstetrics Services for Uncomplicated Cases	0.24	0.43
Obstetrics Services for Uncomplicated and Complicated Cases	0.21	0.41
Obstetrics Services for Serious Illness	0.30	0.46
High Technology Hospital	0.42	0.49
Birthing Room	0.47	0.50
Neonatal ICU	0.14	0.35
Physicians per Bed	0.84	0.48
Nurses per Bed	1.40	0.50
Percentage of Registered Nurses	0.79	0.15

Note: Means are for births or, equivalently, for hospitals weighted by number of sample births.
Source: Lewin analysis of the Birth Product Line Data.

Appendix D: Production Possibility Frontiers and Cost Functions

In this report, we posit that observed relationships between cost and attributes of health outcomes reflect an equilibrium that is determined by the interactions of patients, their families, physicians, other providers, hospital owners, third-party payers, regulators, etc. The function embodies both the available technologies and the preferences and resources of the various actors. In this appendix, we discuss the economics of the production side, and its applicability to the problem of examining multi-dimensional health outcomes. We first consider trade-offs in production (production possibility frontiers). We then consider how the minimum cost of production varies with input prices across given output levels (cost functions). We discuss: how these functions can be used to compare the efficiency in production across producers; approaches to estimating such functions; and their applicability to the problem of understanding observed relationships between cost and health care outcomes.

A. Trade-offs in Production

Technical factors determine rates at which it is technically feasible to trade one component of the burden of disease for another along "production possibility frontiers" (PPFs). PPFs and input prices jointly determine the efficient cost trade-offs between outcome components (Cost Functions). Knowledge of these functions alone is sufficient to compare the productive efficiency of two systems, but this falls short of the desired comparison because it says nothing about how well the system is satisfying consumer needs. Both PPFs and Cost Functions are frontier functions. Economists have used both linear programming and "frontier" econometric models to estimate efficient technical and cost trade-offs between outputs along these frontiers.

1. Production Possibility Frontiers

Economists describe the technological possibilities of a firm or an economy using a mathematical function called the production possibility frontier (PPF). A PPF is the locus of technologically efficient combinations of inputs and outputs given current technology. At every point on this locus, no more of one output can be produced without producing less of at least one other or without obtaining more resources. Symmetrically, no less of any resource can be used without producing less of at least one output or obtaining more of at least one input. A firm or society is said to be *technically* efficient if it is operating at any point on its frontier; i.e., if it is using the minimum combination of inputs to achieve a given level of output, or, vice versa, achieving the maximum level of output with a given level of inputs. Output combinations outside the frontier are not attainable without increasing inputs or improving the technology, and output combinations inside the frontier are technically inefficient because more could be attained with the same inputs.

At every point on the PPF, there is an implicit set of rates at which each pair of outputs can be traded for one another, holding inputs constant. This is called the marginal rate of transformation.

2. Cost Functions

In principle, a firm could use many different combinations of inputs to produce a given combination of outputs in a technically efficient manner. Some, however, will be more expensive than others will. A firm is said to have achieved *allocative* efficiency, if it uses the

minimum cost input combination to achieve any given output combination. Thus, while many possible input combinations may be technically efficient for a given set of outputs, very few (perhaps just one) are efficient allocations.

Which input allocations are efficient for a given set of outputs will depend on prices that the firm must pay for its inputs? Hence, given the technology, input prices and outputs uniquely determine the minimum cost of production.

The cost function gives the minimum cost of producing a given set of outputs with given input prices. Formally, the cost function can be derived by minimizing accounting cost (i.e., the price-weighted sum of inputs) for fixed output levels subject to the technological constraint represented by the PPF. The solution to this problem is a system of derived input demands, in which the demand for each input at the cost minimum is a function of the outputs and the input prices. Substitution of the derived demand equations into the cost accounting equation for cost produces the cost function. Like the derived input demands, the cost function's arguments are outputs and input prices.

The cost function implies an output frontier that is similar to the output frontier of the PPF. The cost function represents the maximum output combinations that can be achieved at given cost and given input prices. Holding input prices, cost and technology constant, outputs beyond the frontier implied by the cost function cannot be attained without increasing cost, while production at any point inside the same frontier is inefficient – more could be produced without increasing cost. The source of inefficiency may be allocative or technical.

The derivative of the cost function with respect to an output is the marginal cost of producing that output. The ratio of the marginal cost for one output relative to that for another at a point on the cost function is the rate at which one output can be traded for the other holding costs constant.

3. Measuring Production Performance

Conceptually, at least, either the PPF or the cost function could be used to compare the performances of firms or systems on the production side. The questions they can be used to address are of the nature:

1. Is System A more efficient technically than System B (based on PPF)? How large is the difference?

2. Are resources allocated more efficiently under System A than under System B (based on the cost function)? How large is the difference?

In the standard textbook treatment of production processes, these questions can be answered without knowledge of who receives the goods and services that are produced. In the context of health care, we might ask whether one system provides specific health services in a more efficient manner than another, or allocates resources used to provide a specific set of services more efficiently than another. But this is a very limited question for two reasons. The first is that the real output of the production process is consumer health, not the health care services per se. Fairly simple extensions of the PPF and cost function concepts can be used to address

this. The second is that ultimately we are interested in whether health outcomes are those desired by consumers, given the cost of achieving them, and whether they are distributed efficiently and equitably among consumers. To address such issues, it is necessary to look beyond trade-offs in production.

Economists have modified the textbook model to accommodate products that depend on characteristics of the consumers who use them, known as "throughputs." Education and health are commonly used examples. In essence, the consumer's characteristics (e.g., health condition) shift the PPF and cost function.

Given a PPF, technical inefficiency for a firm or system can be quantified with measures of the "distance" between the PPF and the actual point of operations (i.e., between actual set of values for the inputs and outputs). The distance between the actual point and a specific point on the PPF might be defined as the square root of the sum of squared differences between the individual coordinates of each point. This distance will depend on which point on the PPF is selected. The minimum distance between the actual point and any point on the PPF might be used, but there are other ways to select a point on the PPF, too (e.g., hold the inputs constant and find the minimum distance among points on the PPF that differ from the point of operation only in the outputs).

Given a cost function, the technical inefficiency of a firm or system can be quantified as the difference between the actual cost and minimum cost of producing a given set of outputs. This difference depends on both the set of outputs and on input prices.

B. Estimation

Two approaches are used in the estimation of production possibility frontiers and cost functions: data envelopment analysis (DEA) and stochastic frontier estimation (SFE). DEA is nonparametric and deterministic; SFE is parametric and stochastic. DEA uses linear programming; the maximum likelihood principal is usually applied to SFE. These two approaches are more complements than substitutes (Kooreman, 1994). We describe each in some detail, then consider their applicability to the analysis of health outcome data for individuals.

C. Data Envelopment Analysis

Consider first the application of DEA to estimation of a PPF. The data are points in a multi-dimensional input/output space. For health data from a specific provider or health system, each point might represent an individual and the data would be an array of health outcomes and inputs used in producing the outcomes. Throughputs – characteristics of the individual upon entry into the system – would also be included.

DEA finds the production possibility frontier by forming an envelope of the data points. This envelope can be thought of as an outer shell of the observed data points. Linear programming is used to search the combinations of inputs and outputs to find the outer shell. This outer envelope is the locus of the most efficient outcomes in the sample. While it is possible that outcomes outside this envelope can be achieved (i.e., the envelope is not the PPF), the

envelope represents the most efficient outcomes that have been achieved by the system. Once this envelope is found, the inefficiency of each interior point can be measured as the distance to the envelope, and the envelope from one health system can be compared to that of another (Folland et al, 1997).

An attractive feature of DEA is that it does not require any assumptions about functional form or error distributions. It is completely nonparametric. Not having to presuppose a specific technology reduces bias due to specification error. This is also a limitation, however. The methodology assumes that all of the "extreme points" are extreme because of efficiency, whereas unmeasured characteristics of the individual, unusual "luck" in treatment, or even measurement error may make these points extreme. Thus, with no error to represent random shocks, if one point in the sample is extreme because of exceptional luck, other points will be viewed as inefficient relative to this point, but falsely so.

DEA can also be applied to the estimation of a cost function. In this case the envelope would be constructed in output-cost-input price-throughput space. If data are from a regime in which prices are constant, they can be ignored – although this will limit the utility of comparing the resulting envelope to that from other systems, where input prices may be different.

D. Stochastic Frontier Estimation

Stochastic Frontier Estimation (SFE) is usually applied to estimation of single-output production functions or to cost functions with multiple outputs. The methodology assumes a single dependent variable, with one or more "exogenous" explanatory variables. In cost function estimation, cost is specified to be a function of outputs or output attributes, and input prices. In an application to health data, throughputs would need to be added, and input prices would drop out if all data points were from a regime with constant input prices.

Unlike DEA, SFE allows for random deviations from the frontier, hence the term stochastic. Costs associated with a case may be affected by unobserved random shocks that are not related to the efficiency of management. Whereas DEA would classify these firms as inefficient, SFE makes a distinction between efficiency and random shocks (luck).

The model is a variant of a multiple regression model (Maddala, 1997):

$y = X\beta + e,$

where:

- y is an n x 1 vector of costs for the n observations (usually expressed in log form);
- X is an n x k matrix of data on k outputs, input prices, and throughputs; and
- e is the disturbance term.

The disturbance term is composed of two parts:

$e = v + u,$

where:

- v is a random disturbance (as in the usual multiple regression model); and
- u is the observation's inefficiency.

The inefficiency component of the disturbance is always nonnegative, reflecting the assumption that each observation lies on or above the cost function. To complete the specification, it is necessary to make assumptions about the distributions of u and v. The most common assumptions are to assume that v is normal and u one of the following: half or truncated normal, exponential, or gamma. The model is estimated via maximum likelihood.

In the applied literature, cost functions are estimated with firm-level data. In this context they are also frequently estimated jointly with derived input demand equations. It seems reasonable to apply the methodology to individual case data in the context of health care, however. Variation in input prices in these data would permit estimation of the derived demand equations as well, provided that data on both input prices and inputs are available.

The main disadvantage of SFE is that the strong assumptions made in the specification of the model – exogenous explanatory variables, a specific functional form, and specific error distributions – may be incorrect and result in biased parameter estimates. As discussed in the text and ***Appendix A***, evidence from the BPL data demonstrates that output – morbidity – is endogenous. That is, random variation in morbidity, after risk adjustment, induces variation in costs. The SFE method relies on the assumptions to separate the inefficiency component of the disturbance from the "luck" component. While this may be problematic, it is perhaps preferred to ignoring the luck component of the disturbance entirely, as DEA implicitly does (Kooreman, 1994).

E. Discussion

While both PPFs and cost functions are useful analytic tools, cost functions are likely to be of more utility in addressing burden of disease issues because they embody information about the efficiency with which resources are allocated, as well as technical efficiency. There are, however, two general approaches to obtaining an empirical cost function. One is to estimate it, directly. The other is to estimate the production possibility frontier and then solve analytically for the cost function (i.e., solve the cost minimization problem described earlier, using the estimated PPF). Simplicity makes the direct estimation of the cost function attractive, but the nature of the data may dictate which approach is preferred.

When data are for individuals, as with the BPL data, the function estimated may depend on whether the data are from patients treated by a single provider, or by many different providers, as well as on the availability of data for cost, inputs, and input prices. Input prices may be considered fixed for a single provider, at least over a reasonably short period, so they could be ignored in the estimation of a cost function. Fixed input prices might also be a reasonable assumption for multiple producers in a metropolitan area, state or even region. When input prices likely vary substantially over sample observations, they should be measured and included as arguments in cost function estimation. Alternatively, one might include dummy

variables to control for various price regimes (e.g., for metropolitan area), but this will not provide information about how changes in input prices affect cost, and the dummies might proxy for other factors that we do not want to control for, including system efficiency. If input data are available, then estimation of the PPF may be preferred. If both input and input price data are available, the cost function may be estimated jointly with the derived input demand functions.

Exogeneity of output is another problematic issue in cost function estimation. That is, the allocation of inputs to the production of a given set of outputs, which determines cost, is independent of any unobserved factors that affect the choice of outputs. In a market economy, unobserved factors generally have an influence on both production costs and output quantities. One of the assumptions in the SFE approach to cost function estimation is that the luck term in the disturbance is independent of the outputs and input prices, and this will be violated if unobserved factors affect both outputs and costs. In the absence of exogeneity, one might specify and estimate a system of equations for simultaneous determination of costs and outputs, but this is a heroic task and is seldom attempted.

In health care, some would argue that outputs really are exogenous to cost because "cost is no object" in treatment decisions. Such stereotypical statements do injustice to reality. In fact, it may be argued that violation of output exogeneity is more extreme in health care than in other goods and services because production and consumption go hand in hand. As already discussed, the consumer (patient) is a throughput in the production function. While we can control for some consumer characteristics, this will be imperfect, and unobserved consumer characteristics can influence both cost and outcomes. Further, the consumer may exert a substantial influence over treatment decisions, again affecting both outcomes and cost. That is, the relationship between cost and outcomes that is observed may not just reflect the production side of the market. Instead, consumer preferences may play a critical role, in which case it might not be possible to consider the production side of the market alone. Finally, and perhaps most importantly, random outcome variation is an important, positive determinant of cost variation, because providers routinely use resources in the treatment of adverse health outcomes.

In short, instead of viewing empirical relationships between costs and outcomes as being determined solely by the production side of the market, it may be more appropriate to view them an "equilibrium" consequence of the interplay between production and consumption factors, reflecting both production trade-offs and consumer preferences.